TREATISE ON BASIC PHILOSOPHY

Volume 1

SEMANTICS I: SENSE AND REFERENCE

TREATISE ON BASIC PHILOSOPHY

1

SEMANTICS I *Sense and Reference*

2

SEMANTICS II *Interpretation and Truth*

3

ONTOLOGY I *The Furniture of the World*

4

ONTOLOGY II *A World of Systems*

5

EPISTEMOLOGY I *The Strategy of Knowing*

6

EPISTEMOLOGY II *Philosophy of Science*

7

ETHICS *The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 1

Semantics I:

SENSE AND REFERENCE

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND / BOSTON-U.S.A.

Library of Congress Catalog Card Number 74–83872

ISBN-13: 978-90-277-0572-3 e-ISBN-13: 978-94-010-9920-2
DOI: 10.1007/978-94-010-9920-2

Published by D. Reidel Publishing Company,
P.O. Box 17, Dordrecht, Holland

Sold and distributed in the U.S.A., Canada, and Mexico
by D. Reidel Publishing Company, Inc.
306 Dartmouth Street, Boston,
Mass. 02116, U.S.A.

All Rights Reserved
Copyright © 1974 by D. Reidel Publishing Company, Dordrecht, Holland
No part of this book may be reproduced in any form, by print, photoprint, microfilm,
or any other means, without written permission from the publisher

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and of right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient – except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: "Do your own thing. Your reward will be doing it, your punishment having done it".

CONTENTS OF *SEMANTICS I*

PREFACE XI

ACKNOWLEDGEMENTS XIII

SPECIAL SYMBOLS XV

INTRODUCTION 1

1. Goal 1
2. Method 4

1. DESIGNATION 8

1. Symbol and Idea 8
    1.1. Language 8
    1.2. Construct 13
    1.3. Predicate 15
    1.4. Theory and Language 18
2. Designation 21
    2.1. Name 21
    2.2. The Designation Function 23
3. Metaphysical Concomitants 26
    3.1. Basic Ontology 26
    3.2. Beyond Platonism and Nominalism 27

2. REFERENCE 32

1. Motivation 32
2. The Reference Relation 34
    2.1. An Unruly Relation 34
    2.2. Immediate and Mediate Reference 36
    2.3. Reference Class 37
    2.4. Factual Reference and Object Variable 39
    2.5. Denotation 42
    2.6. Reference and Evidence 43
    2.7. Misleading Cues in the Search for Referents 46
3. The Reference Functions 48
    3.1. Desiderata 48
    3.2. Principles and Definitions 50

3.3. Some Consequences 53
3.4. Context and Coreference 56
4. Factual Reference 59
4.1. The Factual Reference Class 59
4.2. The Factual Reference Class of Scientific Theories 62
4.3. Spotting the Factual Referents: Genuine and Spurious 68
4.4. The Strife over Realism in the Philosophy of Contemporary Physics 70
5. Relevance 75
5.1. Kinds of Relevance 75
5.2. The Paradox of Confirmation as a Fallacy of Relevance 79
6. Conclusion 81

3. REPRESENTATION 83
1. Conceptual Representation 83
2. The Representation Relation 87
2.1. A Characterization 87
2.2. The Multiplicity of Representations 93
2.3. Transformation Formulas and Equivalent Theories 97
3. Modeling 99
3.1. From Schema to Theory 99
3.2. Problems of Modeling 101
4. Semantic Components of a Scientific Theory 104
4.1. Denotation Rules and Semantic Assumptions 104
4.2. Philosophical Commitment of the SA's 108
4.3. Application to Quantum Mechanics 111
5. Conclusion 113

4. INTENSION 115
1. Form is not Everything 115
1.1. Concepts of Sense 115
1.2. Extension Insufficient 118
1.3. 'Intensional': Neither Pragmatic nor Modal 120
2. A Calculus of Intensions 123
2.1. Desiderata 123
2.2. Principles and Definitions 124
2.3. Main Theorems 126
2.4. Intensional Difference and Family Resemblance 130
3. Some Relatives – Kindred and in Law 134
3.1. Logical Strength 134
3.2. Information 135
3.3. Testability 138
4. Concluding Remarks 140

5. GIST AND CONTENT 142

1. Closed Contexts 143
  1.1. Closed Contexts and Their Structure 143
  1.2. The Logical Ancestry of a Construct 145
2. Sense as Purport or Logical Ancestry 146
  2.1. Purport and Gist 146
  2.2. The Gist of a Basic Construct 148
  2.3. The Gist of a Theory 150
  2.4. Changes in Gist 153
3. Sense as Import or Logical Progeny 154
  3.1. The Logical Progeny of a Construct 154
  3.2. Import 156
  3.3. Theory Content 158
  3.4. Empirical and Factual Content 160
  3.5. Changes in Import and Content 165
4. Full Sense 166
5. Conclusion 171

BIBLIOGRAPHY 173

INDEX OF NAMES 181

INDEX OF SUBJECTS 183

PREFACE TO *SEMANTICS I*

This is a study of the concepts of reference, representation, sense, truth, and their kin. These semantic concepts are prominent in the following sample statements: ⌜The field tensor *refers* to the field⌝, ⌜A field theory *represents* the field it refers to⌝, ⌜The *sense* of the field tensor is sketched by the field equations⌝, and ⌜Experiment indicates that the field theory is *approximately true*⌝.

Ours is, then, a work in philosophical semantics and moreover one centered on the semantics of factual (natural or social) science rather than on the semantics of either pure mathematics or of the natural languages. The semantics of science is, in a nutshell, the study of the symbol-construct-fact triangle whenever the construct of interest belongs to science. Thus conceived our discipline is closer to epistemology than to mathematics, linguistics, or the philosophy of language.

The central aim of this work is to constitute a semantics of science – not any theory but one capable of bringing some clarity to certain burning issues in contemporary science, that can be settled neither by computation nor by measurement. To illustrate: What are the genuine referents of quantum mechanics or of the theory of evolution?, and Which is the best way to endow a mathematical formalism with a precise factual sense and a definite factual reference – quite apart from questions of truth?

A consequence of the restriction of our field of inquiry is that entire topics, such as the theory of quotation marks, the semantics of proper names, the paradoxes of self-reference, the norms of linguistic felicity, and even modal logic have been discarded as irrelevant to our concern. Likewise most model theoretic concepts, notably those of satisfaction, formal truth, and consequence, have been treated cursorily for not being directly relevant to factual science and for being in good hands anyway. We have focused our attention upon the semantic notions that are usually neglected or ill treated, mainly those of factual meaning and factual truth, and have tried to keep close to live science.

The treatment of the various subjects is systematic or nearly so: every basic concept has been the object of a theory, and the various theories have been articulated into a single framework. Some use has been made of certain elementary mathematical ideas, such as those of set, function, lattice, Boolean algebra, ideal, filter, topological space, and metric space. However, these tools are handled in a rather informal way and have been made to serve philosophical research instead of replacing it. (Beware of hollow exactness, for it is the same as exact emptiness.) Moreover the technical slices of the book have been sandwiched between examples and spiced with comments. This layout should make for leisurely reading.

The reader will undoubtedly apply his readermanship to skim and skip as he sees fit. However, unless he wishes to skid he will be well advised to keep in mind the general plan of the book as exhibited by the Table of Contents. In particular he should not become impatient if truth and extension show up late and if analyticity and definite description are found in the periphery. Reasons will be given for such departures from tradition.

This work has been conceived both for independent study and as a textbook for courses and seminars in semantics. It should also be helpful as collateral reading in courses on the foundations, methodology and philosophy of science.

This study is an outcome of seminars taught at the Universidad de Buenos Aires (1958), University of Pennsylvania (1960–61), Universidad Nacional de México (1968), McGill University (1968–69 and 1970–71), and ETH Zürich (1973). The program of the investigation and a preview of some of its results were given at the first conference of the Society for Exact Philosophy (see Bunge, 1972a) and at the XVth World Congress of Philosophy (see Bunge, 1973d).

ACKNOWLEDGEMENTS

It is a pleasure to thank all those who made useful comments and criticisms, whether constructive or destructive, in the classroom or in writing. I thank, in particular, my former students Professors Roger Angel and Charles Castonguay, as well as Messrs Glenn Kessler and Sonmez Soran, and my former research associates Professors Peter Kirschenmann, Hiroshi Kurosaki, Carlos Alberto Lungarzo, Franz Oppacher, and Raimo Tuomela, and my former research assistants Drs David Probst and David Salt. I have also benefited from remarks by Professors Harry Beatty, John Corcoran, Walter Felscher, Joachim Lambek, Scott A. Kleiner, Stelios Negrepontis, Juan A. Nuño, Roberto Torretti, Ilmar Tammelo, and Paul Weingartner. But, since my critics saw only fragments of early drafts, they should not be accused of complicity.

I am also happy to record my deep gratitude to the Canada Council for the Killam grant it awarded this research project and to the John Simon Guggenheim Memorial Foundation for a fellowship during the tenure of which this work was given its final shape. Finally I am grateful to the Aarhus Universitet and the ETH Zürich for their generous hospitality during my sabbatical year 1972–73.

MARIO BUNGE

Foundations and Philosophy of Science Unit,
McGill University

SPECIAL SYMBOLS

C	Set of constructs (concepts, propositions, or theories)
$\mathbb{C}$	Context
$\mathcal{C}$	Content (extralogical import)
$\mathcal{C}n$	Consequence
$\mathcal{D}$	Designation
Δ	Denotation
$\hat{=}$	Representation
$\mathcal{E}$	Extension
$\mathcal{I}$	Intension
$\mathcal{I}mp$	Import (downward sense)
L	Logic
$\mathcal{L}$	Language
$\mathcal{M}$	Meaning
Ω	Universe of objects (of any kind)
$\mathbb{P}$	Family of predicates
$\mathcal{P}ur$	Purport (upward sense)
$\mathcal{R}$	Reference
S	Set of Statements (propositions)
$\mathcal{S}$	Sense
$\mathcal{S}ig$	Signification
T	Theory (hypothetico-deductive system)
$\mathcal{V}$	Truth value function

INTRODUCTION

In this Introduction we shall sketch a profile of our field of inquiry. This is necessary because semantics is too often mistaken for lexicography and therefore dismissed as trivial, while at other times it is disparaged for being concerned with reputedly shady characters such as meaning and allegedly defunct ones like truth. Moreover our special concern, the semantics of science, is a newcomer – at least as a systematic body – and therefore in need of an introduction.

1. Goal

Semantics is the field of inquiry centrally concerned with meaning and truth. It can be empirical or nonempirical. When brought to bear on concrete objects, such as a community of speakers, semantics seeks to answer problems concerning certain linguistic facts – such as disclosing the interpretation code inherent in the language or explaning the speakers' ability or inability to utter and understand new sentences of the language. This kind of semantics will then be both theoretical and experimental: it will be a branch of what used to be called 'behavioral science'. Taking a cue from Chomsky and Miller (1963) we may say that, rather than being a closely knit and autonomous discipline, this kind of semantics is the union of two fields: a chapter of linguistics and one of psychology:

Empirical Semantics
- *Linguistic semantics:* the semantics of natural languages
- *Psycholinguistic semantics:* the psychology of speech acts and contents, or the study of language users

If, on the other hand, semantics concerns itself with conceptual objects, such as mathematical structures or scientific hypotheses, then it remains nonempirical in the sense that it makes no direct use of observation and measurement to test its conjectures and models – nor does it need to, because this kind of semantics does not describe and predict facts. In other words, nonempirical semantics is concerned not only with linguistic items but also, and primarily, with the constructs which some such items stand for as well as with their eventual relation to the real world. (More on constructs in Ch. 1, Sec. 1.2.) This branch of semantics is then closer to the theory of knowledge than the theory of language. (We shall take a look at this point in Ch. 10, Sec. 3.) Moreover, under penalty of being useless, nonempirical semantics should account for our experience with conceptual objects – thus being vicariously empirical. In particular, it should be concerned with our experience in interpreting conceptual symbols, elucidating the sense of constructs, uncovering their referents, and estimating truth values. Furthermore, the way it performs this task should constitute the supreme test of this branch of semantics: there is no justification for a semantic theory that fits neither mathematics nor science nor ordinary knowledge. Conceived in this way, nonempirical semantics may be split up as follows:

- *Nonempirical Semantics*
  - *Basic* or general
  - *Applied* or special
    - S. of mathematics (model theory)
    - S. of science
    - S. of ordinary knowledge

This book is concerned with but a fragment of nonempirical semantics and not at all with empirical semantics. The philosopher as such is not competent to tackle problems in empirical semantics: the best he can do is to study it with the hope of ferreting out the methodology and philosophy of scientific linguistics and psycholinguistics. Of course he may occasionally ask penetrating questions and, with less frequency, propose valuable insights in this field or any other. But, *qua* philosopher, he is not competent to expand such insights into theories proper (i.e. theories in mathematical linguistics), nor is he equipped to design experiments to put such theories to the test. The philosopher, as such, is as much of an amateur in linguistics and psychology as in physics and biology. But he can be a professional philosophical semanticist.

We shall deal exclusively with nonempirical semantics, i.e. with the semantic problems that cannot be investigated by empirical means because they are not concerned with any factual items but, at most, with certain features of our knowledge of such objects. In particular we shall not study the bewildering variety of concepts designated by the ambiguous term 'meaning' (see Schaff, 1962; Cohen, 1966; Hill, 1971). We shall instead restrict ourselves to the semantic concept of meaning, i.e. the sense and reference of predicates, propositions, and theories. In particular we shall not investigate the process by which an organism assigns a significance to a sign: we regard the pragmatic concept of meaning as a business of psycholinguistics (see e.g. Osgood *et al.*, 1957; Luria, 1969), anthropologists, and historians. Likewise we are interested in the semantic concept(s) of truth rather than in the psychological concepts of personal truth, strength of belief, credibility, and so on. Our choice does not imply a rejection of either empirical semantics or pragmatics: it only involves a deliberate restriction of the field of our inquiry – *ergo* a choice of method.

A further restriction on the scope of our investigation will be this. We shall pay special attention to the *semantics of factual science*, one of the three branches of applied or special semantics we recognized a while ago. That is, our inquiry shall focus on the notions of factual reference, factual sense, and factual truth that are relevant to scientific knowledge. The ultimate goal of our research is to accomplish for these notions what model theory has done for the concepts of satisfaction, formal truth, consequence, and extension – without however hoping to attain the neatness characterizing model theory. (For the semantics of logic see van Fraassen (1971), for the semantics of mathematics consult Robinson (1963) or Bell and Slomson (1969).)

In the pursuit of our goal we shall attempt to be both systematic and relevant to real science. To put it in nasty terms: We shall do our best to avoid the two chief blemishes of most of the available views and theories in basic semantics. These shortcomings are the lack of comprehensive systems in which *all* of the semantic concepts hang together and thus illumine one another, and the *irrelevance* to the semantic problems that pop up in live science. We shall try to give a fairly exact treatment but, if we have to choose between a fruitful insight on the one hand, and a rigorous but useless formalism on the other hand, we shall prefer the former. For, as cuckoos and physicists well know, given a fertilized egg there exists a bird willing to hatch it.

Because our semantic system is frankly unorthodox, its measure of success should not be gauged by its agreement with any of the extant views. It should be estimated, rather, by its ability to (*a*) clarify and codify hitherto obscure or stray notions, (*b*) perform suitable semantic analyses of fragments of current factual science, and (*c*) help in the systematic (axiomatic) reconstruction of existing theories in factual science.

2. METHOD

Because our primary concern will be with the semantics of scientific theories, our investigation will rank largely as metatheoretical. Now, every metatheory is couched in a metalanguage allowing us to express certain statements about items occurring in the (object) language used to express the propositions of the object theory. Throughout our quest we shall use as strong a metalanguage as required, for our aim is rather to get a few things done than to economize. We shall employ whatever tool may seem promising to attain precision, unity and clarity: in particular we shall use elementary set theory, a few algebraic theories, and a sprinkling of topology. (For a superb précis of the mathematics needed by the exact philosopher see Hartnett (1963, 1970).) In this sense our semantics will be in line with the semantics of Tarski and Carnap. It differs from theirs (*a*) in its *object*, which is factual science instead of mathematics, (*b*) in its *philosophical underpinnings*, which are realist rather than either nominalist, empiricist, of Fregean, and (*c*) in its *degree of formalization*, which is lower. While the first two differences are clear, the third deserves a short explanation.

Our exposé, though quite formal and systematic, will not be formalized in the metamathematical sense of the word. In particular we shall not specify our metalanguage in advance. We shall not do it partly because any such specification places a restriction, often unsuspected, upon the concepts and propositions that can be expressed in that language. And we do not care for such limitations, particularly when it is not clear which they are. For example, we want to be able to speak of infinite conjunctions and disjunctions, such as law statements, and even of nondenumerable sets of propositions, such as the formulas representing the successive positions of a particle in the course of time. To write off such constructs just because they overflow finitary logic would be unwise, and to try and find out in each case which logical theory justifies them would be going beyond the scope of this work. We take it for granted that, given a useful mathematical or scientific concept, there exists a branch of logic on which it can roost.

As stated above, we assume that logic and mathematics are sufficient to uncover the form and structure of every construct. We assume also that they are necessary to build theories aiming at elucidating and systematizing the semantic concepts of factual reference, factual sense, and factual truth. But, of course, we do not claim that logic, mathematics and the semantical theories contrived with their help will suffice to reveal the syntax and the semantics of every particular scientific construct – any more than geometry suffices to triangulate the universe. However, a combination of logic, mathematics, semantics and substantive knowledge can do the trick of uncovering the formalism and the semantics of a scientific theory. At any rate nothing else has done it, so it is worth trying.

The preceding assumptions concerning the role of logic and mathematics in constructing philosophical theories, such as semantics, characterize what may be called *exact philosophy* (see Bunge, ed., 1973a). They are likely to be challenged by philosophers of a traditional cast of mind. But they may also be doubted by progressives. A case in point is the notion of sense (or intension or connotation or content) – a favorite with conservatives as well as a *bête noire* of progressives for being allegedly impregnable to mathematization. We shall meet the challenge by proposing a mathematical theory of sense – or rather three such theories, one for each component of the sense of a construct (Chs. 4 and 5). Should these particular theories fail, others could take their place: approaches and programs die only if nobody works on them. After all, simpler ideas, like that of many, or class, were regarded as typically nonmathematical, hence unclear, until about one century ago. In any event, a bill of difficulties and failures does not constitute a philosophy.

Ordinary language philosophers and hermeneutic philosophers will complain that our method is misplaced, because logic and mathematics are incapable of discerning the structural subtleties of ordinary language (which one pray?). And if smart they will score a point of two, either because the locution in question has not yet been tamed or because the mathematicians have failed to pay proper attention to it. Given chance and motivation, mathematicians will tackle any problem of form: after all this is their province.

What holds for the analysis of form is also true of the analysis of meaning. Thus at first sight it would seem that 'or' and 'and' are not commutative in ordinary language: that their place makes a difference to meaning. If this were true then logic would be guilty of a gross oversimplification that would render it incompetent to analyze ordinary language. Example:

⌜You should study as well as play.⌝ (1)

does not mean the same as

⌜You should play as well as study.⌝ (2)

These statements are obviously different. Yet their ostensive structure, namely $s \,\&\, p$, and $p \,\&\, s$, is the same. Now, 'ostensive' signifies "superficial". There are hidden assumptions (presuppositions) that must be unearthed. In the case of (1) the presupposition is that the interlocutor, who is supposed to study, is neglecting his studies, whereas in the case of (2) the presupposition is that he is studying and neglecting his playing. More suitable formalizations would therefore be

$$\begin{aligned}(1) &= \neg s \,\&\, p \,\&\, Obl\,(s \,\&\, p)\\ (2) &= s \,\&\, \neg p \,\&\, Obl\,(s \,\&\, p) \neq (1),\end{aligned}$$

where '*Obl*' stands for the deontic operator "ought", taken care of by deontic logic.

Conclusion: As with structure analysis so with meaning analysis: no genuine shade of meaning need remain uncovered if our analysis is pushed far enough. The resulting depth of our analysis will depend on the power of the analytic tools we have employed. What holds for logical form and for meaning holds for philosophical problems in general: if genuine they can be and ought to be tackled in an exact manner. Even though some exact philosophers may lack the *esprit de finesse* characterizing some inexact philosophers, the problems the latter discover can only be solved with a pinch of *esprit de géométrie*.

A final remark on method. We shall be concerned throughout with theories and their components, i.e. statements, or the designata of declarative sentences. Hence we need not investigate nonalethic constructs such as problems (expressed by questions or by commands) and norms (expressed by imperatives). In principle these and other nondeclarative phrases may be treated by either of two methods: the direct and the indirect. The *direct* procedure consists in taking the bull by the horns and building theories (e.g. systems of deontic logic and of erotetic logic) legalizing, codifying and so cleansing some of our naive ideas on the subject. The *indirect* method consists in transforming the problem by translating the given nondeclarative sentence into a declarative one – i.e. in shedding the pragmatic trimmings of the original phrase. For example, 'Run!' may be transformed into 'The subject is ordered to run', and 'Where is x?' into 'The question is to locate x'. The problem whether every nondeclarative sentence has a "declarative prototype" (Marhenke, 1950) or a "propositional content" (Searle, 1969) seems open but does not affect our enterprise. We need not commit ourselves on this matter and we need not espouse either method for, as pointed out a while ago, scientific theories – our chief analysanda – contain just statements: only the research process leading to and from theories involves questions, norms, promises, threats, etc. However, our results concerning reference, sense, and truth, can be applied to nonalethic constructs provided the latter can be translated or distilled into their alethic counterparts. In this way we can speak of the meaning of problems and norms, or of the significance of questions and commands. In other words, we shall adopt the following principle: *If a construct has an alethic counterpart then the semantics of the former (though not its pragmatics) equals the semantics of the latter.*

Having delineated the goal and method of our endeavor, we proceed to take off.

CHAPTER 1

DESIGNATION

The goal of this chapter is to characterize the most basic of all semantic concepts: that of designation. This concept occurs in statements such as ⌜Sign x designates concept (or proposition) y⌝. Since a significant sign is a member of some language or other, we must start by defining the notion of a language and, in particular, that of a conceptual language, i.e. one capable of expressing propositions. But this will necessitate a clarification of the very nature and status of concepts and propositions. Which will in turn lead us to discussing some of the philosophical underpinnings and ramifications of our enterprise – for philosophical semantics, though a distinct discipline, is not an isolated one.

1. SYMBOL AND IDEA

1.1. *Language*

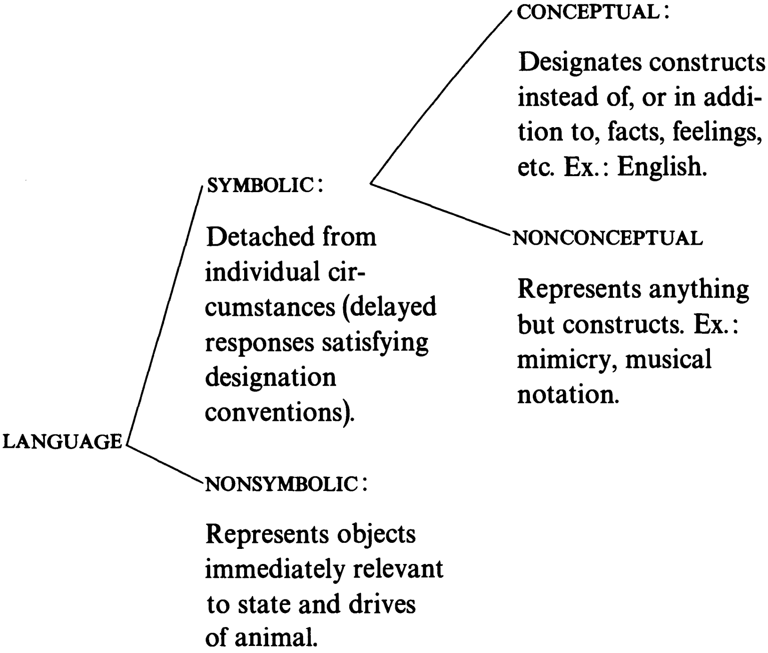
An artificial sign, whether written, uttered, or in any other guise, is a physical object – a thing or a process that a thing undergoes. But of course it is a very special object, namely one that

(i) *represents* some other object (physical or conceptual) or is part of some much proxy,

(ii) belongs to some *sign system* (= language), within which it can concatenate with other signs to produce further signs, and such that the whole system is used for

(iii) the *communication* or transmission of information concerning states of affairs, ideas, etc.

For something to be called a sign it need not be spoken or written: the signs used by bees and apes satisfy all the preceding conditions. Consequently a language need not be symbolic (i.e. involving conventions), let alone conceptual. Any *system of coded signals used for communication purposes* qualifies as a language. Moreover, the coding and decoding need not be accompanied by understanding: they can be automatic – as is often the case within humans. In any event we have the following possible partition of languages.



LANGUAGE
- SYMBOLIC: Detached from individual circumstances (delayed responses satisfying designation conventions).
  - CONCEPTUAL: Designates constructs instead of, or in addition to, facts, feelings, etc. Ex.: English.
  - NONCONCEPTUAL Represents anything but constructs. Ex.: mimicry, musical notation.
- NONSYMBOLIC: Represents objects immediately relevant to state and drives of animal.

We are interested only in symbolic languages: the study of our various systems of growls, snarls, shrieks, and gestures belongs to psychology. (Moreover we are not interested in individual symbols but rather in equivalence classes of such: linguistics is concerned with the tee shape rather than with any concrete instance of the sound or letter tee.) Essentially, a symbolic language is a set of basic symbols (the alphabet) that can concatenate forming strings. A formation device picks the subset of expressions or well formed formulas, which in turn designate certain objects. And a transformation device converts certain expressions into others. More precisely, we shall adopt the following characterization of the overall structure of a finitary symbolic language (whether or not conceptual):

DEFINITION 1.1 A septuple $\mathscr{L}_K=\langle\Sigma, \Box, \circ, \phi, \tau, \Omega, \Delta\rangle$, where K, Σ and Ω are sets, $\Box$ is a distinguished element of Σ, $\circ$ is an operation on Σ, ϕ and τ are families of mappings, and Δ is a function, is called a (finitary) *symbolic language* for communicating systems of kind K iff

(i) Σ, the *alphabet* of $\mathscr{L}_K$, is denumerable and every element of Σ is a symbol that any member of the class K of things can produce or receive;

(ii) the structure $\langle\Sigma^*, \Box, \circ\rangle$, where Σ^* is the set of finite concatenations (*strings*) of elements of Σ, is the free monoid generated by Σ, with identity (neutral) element $\Box$ (the *blank*);

(iii) ϕ, the *formation* device, is a collection of maps from the n-tuples of strings in $(\Sigma^*)^n$ to a subset Σ^{**} of Σ^* (the *expressions*, or *complete utterances*, or *wff*'s of $\mathscr{L}_K$);

(iv) τ, the *transformation* device, is a collection of maps from the n-tuples of wff's in $(\Sigma^{**})^n$ into Σ^{**};

(v) every element of the set Ω of *objects* is associated with (or it evokes) a definite state in an arbitrary member of the set K of users of $\mathscr{L}_K$ (but not conversely: some states of all users correspond to no members of Ω);

(vi) Δ, the *coding* function (or *interpretation* map) of $\mathscr{L}_K$, is a many-one function from the expressions in Σ^{**} to the family $\mathscr{P}(\Omega)$ of all the subsets of Ω (i.e. Δ assigns every complete utterance, e.g., a phrase, a certain class of objects).

More explicitly, the membership of Σ is the totality of basic (atomic) signals or letters, including the nonsignal $\Box$. These signals concatenate associatively to form the set of strings (both well formed and ill formed) Σ^*. For example, if a and b are in Σ, then $a\circ(b\circ\Box)=a\circ b$, which may not be the same as $b\circ a$, is a member of Σ^* though not necessarily of Σ^{**}. The maps in ϕ pick those strings (finite concatenations) that count as correct expressions (or well formed formulas) of $\mathscr{L}_K$, i.e. the elements of Σ^{**}. For example, one of the functions in ϕ will be that which carries the couple $\langle(x), Px\rangle$, belonging to $(\Sigma^*)^2$, into the expression $(x)\,Px$, that belongs to Σ^{**}. (Because Σ^* is denumerable, though infinite, so is Σ^{**}. In other words, the language we are describing contains only a countable infinity of sentences. This is a serious limitation but it need not concern us here, as it will have no consequences for our future work.) So far the basic ideas of the *syntax* of a language. (For mathematical linguistics see e.g. Chomsky and Miller (1963), Chomsky (1963), Ginsburg (1966), Marcus (1967), Arbib (1968), or Harris (1968), each with a different standpoint.) These basic ideas allow one to define a number of useful derivative notions, among others the following two.

DEFINITION 1.2 Let Σ^* be the set of strings of a language $\mathscr{L}_K$. If x and y are in Σ^*, then x is a *part* of y iff there are two strings w and z in Σ^* such that $w \circ x \circ z = y$.

This part-whole relation is reflexive, antisymmetric, and transitive as it should be. Should x be an initial (or final) segment of y, then we should take w (or z) to be the blank $\square$. We shall need this concept in Sec. 2.1.

DEFINITION 1.3 Let Σ^* be the set of strings of a symbolic language $\mathscr{L}_K$. The *length* of a member x of Σ^* is the number of basic signs (counting repetitions) that are a part of x.

(Note the difference between the length of an expression and the complexity of the corresponding construct if any. Any given construct may be expressed by a number of strings of varying length. More on the difference between constructs and their linguistic wrappings in Sec. 1.2.)

So much, or rather little, for the syntax of $\mathscr{L}_K$. Its *semantics* is given by adding to its syntax a set Ω of objects and a coding function Δ. This function associates every expression in Σ^{**} with a collection of objects (perhaps just a singleton) called the *denotatum* of the corresponding expression. For example, in arithmetic we have, *inter alia*, Δ: Numerals$\to$ Numbers. The function Δ is many-one: any collection of objects may be called by different names. The objects in question may be physical (occasionally other signs) or conceptual. A particular language is specified by determining not only its syntax (essentially its vocabulary Σ and its grammar $\gamma = \langle \phi, \tau \rangle$) but also its semantics – essentially its denotata Ω and its coding function Δ. While the former decrees what strings are well formed, the latter decides which are well informed.

Caution: the *semantics of a language* is sometimes mistaken for a semantic *theory* (e.g. by Katz and Fodor, 1963). The former is an integral part of a language: it consists roughly of its dictionary, i.e. of the coding function $\Delta : \Sigma^{**} \to \mathscr{P}(\Omega)$. On the other hand a semantic *theory* is a hypothetico-deductive system couched in some language and concerned with explicating semantic concepts. In other words, while the semantics of a language boils down to its dictionary, a semantic theory may include a theory about dictionaries but is not supposed to include any particular dictionary. Likewise a theory of chemical binding does not contain a list of chemical compounds but enables one to explain and predict their formation. (For further criticisms of the Katz-Fodor view see Bar-Hillel (1970).) Enough of this because there are no viable semantic theories of natural languages anyway.

It has become fashionable to include, in the semantics of a language, the conditions under which certain expressions of the language (its sentences) are true, i.e., their *truth conditions*. We shall not adopt this policy for the following reasons. First, because we shall assign truth values to (some) propositions rather than to their linguistic expressions (sentences). Second and more important, because the proposing and discussing of truth conditions or criteria is a task for the special sciences not for a general theory of language. A language should be rich and neutral enough to allow one to express in it any number of mutually incompatible truth conditions. What semantics can do is to study the general concept of a truth condition – provided it does it in the light of the experience gained in the sciences instead of legislating a priori. (More on this in Ch. 8, Sec. 2.4.) So much for the semantics of a language.

A more complete specification of any language $\mathscr{L}_K$ should include its *pragmatics*. The pragmatics of $\mathscr{L}_K$ may be construed as a map Π_K of the set Σ^* of expressions (both sound and lame) into a set of behavior items of the members of K, the users of $\mathscr{L}_K$. But since the effective determination of such a function Π_K is a matter for empirical investigation we shall not be concerned with pragmatics here. (True, there have been attempts to build a priori systems of pragmatics, by abstracting from concrete linguistic circumstances – in particular from the psychical and social make-up of the language users. But such attempts are methodologically as misguided as any other aprioristic approach to empirical matters.) Another reason for leaving pragmatics aside is that, *pace* Putnam (1970), pragmatics presupposes semantics in the sense that before attempting to investigate what person x means by expression y, or how person z uses the truth concept, one should be reasonably clear about the semantic concepts of meaning and truth – just as the physicist makes sure he understands the concept of atomic weight before proceeding to measuring atomic weights. More on this in Ch. 10, Sec. 3.4.

We conclude here our hasty characterization of the concept of a language. We take language for granted and leave it in competent hands – those of linguists, psycholinguists, sociolinguists and historical linguists. Our concern is with what can be said with languages of a special kind, namely the conceptual ones. That is, we are interested in following the Δ arrow.

1.2. *Construct*

From now on we shall restrict our attention to *conceptual languages*, which are the ones employed in mathematics, science, and philosophy. In anticipation of the formal definition to be given in Sec. 2.2 we may say that what characterizes a conceptual language is that some of its expressions symbolize ideas. If we abstract from ideation, which is a concrete brain process, as well as from communication, which is a concrete physical process, we get *constructs*: concepts (in particular predicates), propositions, and bodies of such – for example, theories. Unlike cognitive psychology, psycholinguistics, and pragmatics, all of which are concerned with real people engaged in thinking and communicating, philosophical semantics abstracts from people, and hence does not handle communication. (Nor does mathematical linguistics for that matter.) Philosophical semantics handles constructs *as if* they were autonomous – i.e., Platonic Ideas – without however assuming that there *are* such.

The existence of constructs may be regarded as a pretense on a par with the infinite plane wave and the self sufficient steppenwolf: all three are fictions. (See Vaihinger, 1920; Henkin, 1953.) In each case the real thing is far more complex but, if we want to theorize, we must start by building more or less sketchy models: once the modeling is under way we may contemplate complication and articulation. In any event we shall agree that a conceptual language is a language suitable for expressing constructs of some kind, e.g., biological theories. Moreover we assume that the chief linguistic categories *correspond* to (but are not identical with) the conceptual categories: (some) terms correspond to concepts, (some) sentences to propositions, certain fragments of languages to theories. (Cf. Kneale, 1972.) See Table 1.1., which summarizes the preceding informal characterization of the sign-construct relations.

(Note that we make no use of the notion of a semantic category [*Bedeutungskategorie*] introduced by E. Husserl and worked out by S.

Leśniewski and K. Ajdukiewicz. It does not improve things for us because a semantic category, such as a name or a sentence, is supposed to be "defined by its meaning" (Ajdukiewicz, 1935), and none of those authors had a theory of meaning to offer. In any case their semantic categories are actually syntactic categories, and the elucidation of this concept is a problem for theoretical linguistics not for philosophical semantics. Here we take languages and linguistics for granted.)

TABLE 1.1

Linguistic categories and conceptual categories

Linguistic category		Conceptual category	
		Variable	Constant
Term	Individual name	Unspecified individual x. Paradigm: arbitrary member of a set.	Specific individual c, or individual concept. Paradigm: "3"
	Class name	Unspecified set X. Paradigm: abstract set.	Specific ("concrete") set. Ex.: the set of electrons.
	Predicate letter	Unspecified predicate X. Paradigm: arbitrary predicate.	Specific predicate P. Ex.: a concept of pressure.
Expression	Phrase	Definite description schema. Ex.: "The cube of n".	Definite description. Ex.: "The cube of 2".
	Sentence	Statement schema (propositional function). Paradigms: Px, Xa, Xy.	Specific statement. Ex.: "Wolves do not wage wars".
	Language	Abstract theory. Paradigm: Boolean algebra.	Specific theory (= theory of a model). Paradigm: arithmetic.

The unit construct is the *concept*. The concept of a concept cannot be defined without circularity but it may be characterized in various ways. From a logical point of view concepts are the building blocks of statements. (Or the other way around: propositions are certain compounds of concepts.) *Example 1* All three components of the statement $\ulcorner 3>2 \urcorner$ are concepts. *Example 2* Every one of the components (coordinates) of the relational structure "$\langle N, +, 0\rangle$", where 'N' designates the set of natural numbers, is a concept and so is the structure as a whole. *Example 3* Every symbol in Newton's formula for the gravitational force designates a concept, and the formula as a whole designates a proposition (which in turn represents an objective stable pattern).

We may distinguish two kinds of concept: individual, such as "Mars", and collective, such as "planet". The latter, i.e. class concepts, are usually called *predicates.* Predicates can be unary like "long", binary like "longer than", ternary like that occurring in "a is b times longer than c", and so on. Let us for a moment focus our attention upon predicates.

1.3. *Predicate*

We shall analyze the concept of a predicate with the help of the mathematical concept of a function. A *function* f is a correspondence between two sets A and B such that, to every member x of A, there is a single element y of B. The correspondence is written '$f:A \rightarrow B$', where A is called the *domain* and B the *range* of f. The value f takes at $x \in A$ is designated by $f(x)$, in turn an element y of B. That is, $f(x) = y$. This is the gist of the general concept of a function.

We are interested in a particular kind of functions, namely propositional functions. A *propositional function* P is a function whose values are propositions. That is, a propositional function, or *predicate*, is a function from individuals to statements. Thus "lives" ("is alive") may be regarded as a certain mapping L from a set D of objects such that, for an individual c in D, $L(c)$ is the proposition "c is alive". Briefly, $L:D \rightarrow S$, where the domain D is in this case the set of organisms and S a certain set of statements, namely the class of propositions in which the concept L occurs. Likewise "dissolves" may be analyzed as a function from the set of ordered pairs ⟨solvent, solute⟩ to a set of statements. In general, a predicate (or propositional function) of *rank* (or *order*) n, where n is some integer greater than zero, will be analyzed as a function

$$P: A_1 \times A_2 \times \cdots \times A_n \rightarrow S$$

where each A_i, for $1 \leqslant i \leqslant n$, is a set of objects, S is a class of statements (propositions), and the cross represents the cartesian product of the sets of objects concerned. In simpler terms: a predicate of rank n is that which combines n objects, say $x_1, x_2, \ldots, x_n$, not necessarily real nor necessarily distinct from one another, to produce something else, namely $Px_1x_2 \ldots x_n$, called a *statement* or *proposition*, and constituting the value of P at the point $\langle x_1, x_2, \ldots, x_n\rangle$. Moreover, given an arbitrary function $f: A \to B$ the corresponding predicate will be taken to be the propositional function

$$P: A \times B \to S \quad \text{such that} \quad Pxy = (f(x) = y) \quad \text{for} \quad x \in A, y \in B.$$

The previous analysis of a predicate as a function applies to atomic, i.e. logically simple predicates, such as "between". Because we have construed atomic predicates as functions we must build complex (molecular) predicates by respecting the formation rules of complex functions out of simpler functions. Thus just as the sum and the product of two functions are definable only on their common domain, so *the disjunction and the conjunction of two predicates must be defined on their overlap* – provided the latter is not void. Otherwise the predicate letter will fail to stand for a genuine predicate: it will be a nonsense sign. In short, we stipulate that, if P and Q are predicates with a common domain $D = A_1 \times A_2 \times \cdots \times A_n$, then

$$\begin{aligned} \neg P: D \to S \quad &\text{such that} \quad (\neg P)\, x_1 x_2 \ldots x_n = \neg (P x_1 x_2 \ldots x_n) \\ P \circ Q: D \to S \quad &\text{such that} \quad (P \circ Q)\, x_1 x_2 \ldots x_n = \\ &\qquad = P x_1 x_2 \ldots x_n \circ Q x_1 x_2 \ldots x_n \end{aligned}$$

where $x_i \in A_i$ for $1 \leqslant i \leqslant n$ and $\circ$ is an arbitrary binary connective, such as "&" or "⇔". The advantages of this construal are multiple. First, it holds for all predicates. Second, it exhibits the referents of a statement. Third, it does not require the concept of truth – whence it is independent of any particular theory of truth. Fourth, it writes off without further ado, as ill formed, compounds such as 'black thought' and 'cleverly melting at 100 °K' because their components, though *bona fide* predicates, are defined on disjoint domains. Such pseudopredicates can be assigned neither sense nor reference.

Note that our construal of a predicate differs from Frege's analysis as a function from individuals to truth values. (Recall Frege (1891), in Angelelli (1967) p. 133: "*ein Begriff ist eine Funktion, deren Wert immer ein Wahrheitswert ist*".) To put it in contemporary mathematical jargon, Frege identifies a predicate F with the characteristic function χ_D of the domain D of F. In short he sets $F = \chi_D: D \to \{0, 1\}$. But then he is unable to distinguish among the various predicates with the same domain, because there is only one characteristic function for each set. Frege's construal of a predicate is therefore unacceptable. Besides, it is inconsistent with his own (wavering) anti-extensionalism and it involves an unanalyzed concept of truth. (More in Ch. 8, Sec. 3.6.)

On the other hand we accept Frege's conception of a proposition (which he often called *Gedanke*, i.e. thought), namely as the designatum of a declarative sentence independent of its particular wording. (This is a rough characterization not a definition.) That was also the manner Bolzano (1837), the Russell of *PM*, and Church thought of propositions as distinct from their linguistic containers (see, e.g. Church, 1956). We shall not define the concept of a proposition (or statement) but rather the whole structure of a metric Boolean algebra of propositions – not until Ch. 8, Sec. 3.2, though. For the time being we shall clarify what we do *not* mean by a predicate and its values (propositions).

The vain attempt to frame quick definitions of "concept" and "proposition" has produced a number of more or less interesting mistakes. *First:* "A concept is the designatum of a grammatical predicate or predicate letter". Counterexamples: the artificial predicate signs 'crittle' and 'analytic or hot' stand for no concepts. Only the converse is true if qualified: every genuine concept can be either named or described by at least one string in some language. *Second:* "A concept is anything that has a sense (*Begriffsinhalt*)". Or, which is nearly equivalent: "anything which is capable of being the sense of a name *x* is called a *concept of x*" (Church, 1951, p. 11). Unnecessary: there are unnamed (though describable) concepts – witness the silent majority of real numbers and functions. And insufficient: propositions and theories too have a sense. *Third*: "A concept is anything that can be assigned a referent (*Begriffsgegenstand*)". Not necessarily so: sets are concepts and yet they do not refer to anything. *Fourth*: "A proposition is the designatum of a sentence". Close but not quite: the nonsense sentences in nursery rhymes and in some philosophical writings express no propositions. *Fifth*: "A proposition is whatever is either true or false". Not quite: an untested factual statement has no truth value (see Ch. 8). To say that it has got one, only we ignore which, is sheer Platonizing and it does not advance us: the truth of the matter of factual truth is that truth values are contingent upon tests. (More in Ch. 8.) *Sixth*: "A proposition is anything that satisfies the propositional calculus". Necessary but not sufficient: other objects besides propositions obey the same algebra. *Seventh*: "A proposition is a collection of synonymous sentences in a language (i.e., an equivalence class of sentences under the relation of synonymy in a language)". Tempting but hollow, for it does not seem possible to give a purely linguistic account of synonymy. In fact we recognize two sentences as having the same significance, over and above their linguistic disparities, just in case they happen to stand for the same proposition. Example: the different sentences 'p & q' and 'q & p' designate the same proposition. In other words equisignificance, a linguistic property, ir reducible to a semantic property, namely the identity of the underlying constructs. (More precisely, as will be argued in Ch. 7, two constructs are identical just in case they have the same meaning, which happens if and only if they have the same sense and point to the same referents.)

So much for a preliminary clarification of "concept" and "proposition". We turn next to theory in relation to language

1.4. *Theory and Language*

The third category of construct is the theory, or hypothetico-deductive system. The formulas of a theory may be specific statements or they may contain unbound variables of some kind. In either case the formulas are formulated in, or expressed by, sentences in some language – the *language of the theory*. One and the same language may be used to formulate any number of alternative theories: when formulating a theory we pick a subset of all the possible expressions of a language and organize this collection (or rather the corresponding set of statements) in a deductive manner. The language itself must be neutral with respect to both the selection and the organization of the material (recall Sec. 1.1). Thus while the language may contain the sentences 'a is a P' and 'a is not a P', the theory will pick either or none of them.

The following example should bear out our thesis that theory and language must be kept as distinct as gift and wrapping. Let $\mathscr{L}$ be the following language:

Logical alphabet of $\mathscr{L} = \{\neg, \vee, \exists\}$

Extralogical alphabet of $\mathscr{L} = \{a, b, x; P\}$ with a and b individual constants, x and individual variable, and P a unary predicate letter

Atomic sentences of $\mathscr{L} = \{Pa, Pb\}$

Sentences of $\mathscr{L} = \{Pa, Pb, \neg Pa, \neg Pb, Pa \vee Pb, \neg Pa \vee Pb, Pa \vee \neg Pb,$
$\neg Pa \vee \neg Pb, (\exists x)Px, \neg(\exists x)Px, (\exists x)\neg Px,$
$\neg(\exists x)\neg Px, \ldots\}$

Theory 1		*Theory* 2
Language of T_1	$=\mathscr{L}=$	Language of T_2
Logic of T_1 =	First order predicate calculus	= Logic of T_2
Axiom 1 $\neg Pa \vee Pb$.		Axiom $\neg(Pa \vee Pb)$
Axiom 2 Pa		Theorem 1 $\neg Pa \wedge \neg Pb$.
Theorem 1 Pb		Theorem 2 $\neg Pa$
Theorem 2 $(\exists x)\, Px$		Theorem 3 $\neg Pb$
		Theorem 4 $(\exists x) \neg Px$.

Although the set of sentences of each theory is a subset of the collection of sentences of their common language, the two subsets do not coincide and moreover they do not stand in their own right but "say" something even if this something is abstract for lack of interpretation of the various symbols involved. Whether a system of signs constitutes a *language* (and is consequently neutral) or *expresses* a theory (and is consequently committed) can be decided only by finding out whether it makes a choice among all the possible sentences – i.e. whether it excludes some of the formulas of the given language.

Nevertheless the language/theory division, though genuine, is relative. Indeed every universal theory, whether in logic or in pure mathematics, can be *used as a language* by another, more specific theory. Thus logic is used as a language or vehicle of communication by every mathematical theory, and mathematics is a language of theoretical science. There are two ways in which a theory can be utilized as a language by another theory. One is to borrow only some of the *concepts* of the universal theory and their corresponding symbols without using any of the axioms and theorems of that theory. This is how most mathematicians use logic. It is also the way molecular biologists use information theory: although they talk of the information carried by a DNA molecule they never compute or measure that quantity of information. Another, fuller, utilization of a theory by a second theory occurs when the latter employs some of the *statements* (hence also some of the concepts) of the former. This is how physicists use functional analysis and sociologists graph theory: these theories enter into the very building of the specific factual theories. In sum, the distinction between language and theory, though neat, is as relative as the one between means and goals.

However a relative difference *is* a difference. Hence it is mistaken to define a theory as a certain set of *sentences* from some language, and a model as a structure in which such strings, conceived as *inscriptions*, are satisfied. Such an elimination of constructs in favor of their linguistic embodiments is no mere carelessness but deliberate: it is a necessary component of nominalism. This philosophy, espoused by eminent logicians such as Hilbert and Tarski, and at one time Quine, has its advantages: it simplifies, it avoids the pitfalls of both Platonism and psychologism and last, but not least, any computer can be talked into adopting it.

However, nominalism does not supply an adequate semantic theory. For one thing it overrates the importance of notation and wording and therefore cannot account for the fact that any given proposition can be formulated in a number of different languages. For another, if one refuses to accept anything beyond signs *qua* physical objects and their eventual physical denotata, then he cannot explain why symbols enter into non-physical (e.g. logical) relations – not to speak of their form and content, which are likewise non-physical. (Consider: whereas twice 3 is 6, twice '3' is '33'.) Thirdly, for the same reason the Ockhamist is likely to multiply the number of entities without necessity, as he may take seriously Eddington's parable of the two tables: the layman's (concept of a) smooth table and the scientist's (concept of a) mostly hollow one, both denoted by a single word. The nominalist may thus end up by positing a plurality of realities, thus committing philosophical suicide – as Chwistek (1949) did. Fourthly, if the nominalist assigns to some of his marks, e.g. sentences, certain non-physical properties such as truth or falsity (as Tarski does), then he slips into hylemorphism – the very Platonic devil he wishes to exorcise. Fifthly, the nominalist must repudiate any theory which, like non-standard analysis, teems with nameless constructs (Robinson, 1966). (For further criticisms see Frege (1893) in Geach and Black, Eds. (1952), and Putnam (1971).)

We shall consequently stress the traditional distinction between term and concept, predicate letter and predicate, sentence and proposition, and language and theory. We shall say that the first member of each pair stands for, expresses, or designates the second member. But this relation of designation deserves a separate section.

2. DESIGNATION

2.1. *Name*

Names are those terms in a language that designate objects of some kind. Thus the numerals '3' and 'III' name the number three. In a conceptual language all names designate constructs. But the converse is false: notwithstanding nominalism, most constructs go unnamed. Thus, only a few irrational numbers and a few functions have got standard names. Still, just as anonymous persons can be identified by their features and actions, so constructs can be identified by their properties even if they are assigned no individual names. For example, the function "two fifths of the cube plus seven" has got no special name but it is unambiguously characterized by the preceding description, which amounts to saying what the function does – namely sending x into $\frac{2}{5}x^3+7$, where x is a real number. In short, there are unnamed constructs but not ineffable ones.

There are names of (some) individuals, names of (some) classes (or common names), of (some) relations, and so on. In either case names may designate a fixed or an arbitrary member of a collection, i.e. an unspecified individual. Thus we may agree to call 'x' an arbitrary real number. (Note that a variable is not quite the same as a blank. While all blanks or voids are the same, namely nothing, variables can differ from one another and can be manipulated. Thus '$x+y=z$' is not the same as ' $+$ $=$ '. Nor is a variable something that varies.)

Names are symbols and as such they act as proxies for their nominata. We should not forget what they stand for – unless we happen to be computers. For example, strictly speaking we should not say 'Let R *be* the real line' but rather 'Let R *name* (*designate*) the real line'. However, as long as we are alert to the distinction between signs and their designata we can afford to confuse them in our speech and writing for the sake of brevity. In short, we can indulge in self-designation provided we keep in mind that symbols are just that – conventional signs proxying for something else.

The refusal to distinguish symbols from what they symbolize can be made into a philosophy. It is, indeed, the core of the nominalist or vulgar materialist philosophy of mathematics. This philosophy appeals to the haters of intangibles and the lovers of simplicity: instead of having symbols on the one hand and constructs on the other it offers a single bag of tangible entities, some natural and others (the signs) artificial. Thus a member of this persuasion will claim, for example, that "The expression or string consisting of '(x) ('followed by 'P' followed by '$x\Rightarrow$' followed by 'P' followed by 'x)' is an analytic sentence of the language L'. And he will be happy with the nursery school identification of the number one with a vertical stroke. A simple creed indeed – hence one inadequate to tackle the complex facts of life. For example the creed makes no room for the principle that symbols, in particular names, are replaceable because conventional: that "the real thing" is the designatum not its name: that no particular sign is indispensable. (For a vigorous defense of this thesis see Frege (1895). Even Bourbaki (1970, Ch. I), despite its nominalist flare-ups, warns against confusing symbol with designatum.)

The preceding semantic principle, that although names may be necessary (or convenient) none of them is indispensable, has been credited to Shakespeare. In fact, in *Romeo and Juliet* he held that the scent of a rose is name-invariant. This principle, sometimes called the "principle of reference" (Linsky, 1967), can be given a more exact though far less poetic formulation, such as the following. Let x be part of an expression $e(x)$ in a language $\mathcal{L}$ and let $e(y)$ be the expression resulting from trading x for y in $e(x)$, where y is yet another sign in $\mathcal{L}$. If x and y have the same designatum, then so have $e(x)$ and $e(y)$. Consequently the two can be mutually substituted *salva significatione et salva veritate*. One may adopt this necessary triviality as part of the definition of the designation relation.

Thus we accept Shakespeare's thesis that nothing is in a name: that what really matters is the nominatum. If the latter happens to be a construct, not a physical object, then we may say, following Frege and Church, that the name points to the sense *or* to the referent of the construct. But we may as well substitute 'and' for 'or', as the concept of ambiguity has a pragmatic ring about it. In our theory of meaning a sign that stands for a concept performs both functions: it signifies the sense as well as the referent of the construct it designates. (See Ch. 7, Sec. 1.) Thus the class name *Homo sapiens* symbolizes the technical concept of man. The sense of this concept is given by some of the hypotheses occurring in the sciences of man, while its reference class is of course the set of humans. Whether or not this latter set is empty (as some of us are beginning to suspect) is no business of the semanticist: the determination of the actual extension of concepts is a task of scientists not of philosophers. But this matter will be taken up in detail in Ch. 2 and in Ch. 9, Sec. 1.

Names are conventional but they need not be arbitrary. While some are formed spontaneously others are contrived according to rule. About the most sophisticated of all rule-directed naming procedures is Gödel's, which assigns to every basic symbol in the linguistic wrapping of a theory a unique natural number in such a way that, given any natural number, the corresponding symbol can be effectively retrieved. We shall need no such powerful method: the most modest of all naming procedures will suffice us – namely the *quotation marks* technique. We shall employ three kinds of quotes:

> 'single' quotes to mention symbols,
> "double" quotes to designate constructs, and
> ⌜corners⌝ to name propositions (a kind of construct).

2.2. *The Designation Function*

According to Definition 1.1 (Sec. 1.1), a language is a certain septuple $\mathscr{L}_K = \langle \Sigma, \square, \circ, \phi, \tau, \Omega, \Delta \rangle$, where Δ is a correspondence between the signs of $\mathscr{L}_K$ (formed from items in Σ with the help of the operation $\circ$) and the objects in Ω. The role of Δ is then to inject some significance into a bunch of dumb marks, which would otherwise symbolize nothing. In the case of conceptual languages, which are the ones we are interested in, the set Ω shrinks to a subset C, a collection of constructs, and Δ becomes the designation function, briefly $\mathscr{D}$. More precisely, we have

DEFINITION 1.4 Let L be a system of predicate logic. Then the septuple $\mathscr{L}_{KL} = \langle \Sigma, \square, \circ, \phi, \tau, C, \mathscr{D} \rangle$ is a *conceptual language* iff

(i) $\mathscr{L}_{KL}$ is a language;

(ii) Σ contains at least the signs required to designate the basic concepts of the logical system L;

(iii) C is a nonempty set of constructs containing at least all those in L;

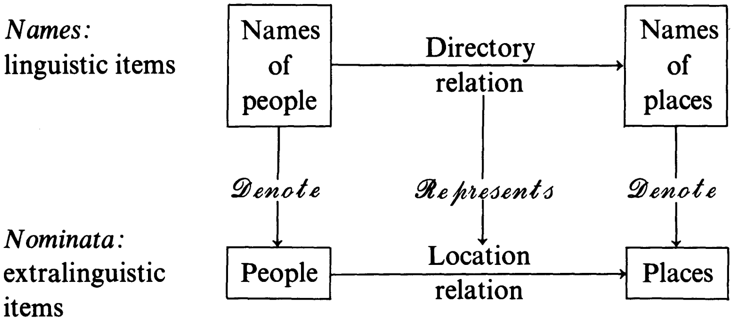
(iv) $\mathscr{D}$ is a many-one function from the set Σ^{**} of expressions (wff's) of $\mathscr{L}_{KL}$ into (but not onto) the collection $\mathscr{P}(C)$ of subsets of C;

(v) Any two signs in Σ^{**} with the same designatum (i.e. for which $\mathscr{D}$ takes the same value) can be mutually exchanged wherever they occur in Σ^{**}.

Note the following points. First, since $\mathscr{L}_{KL}$ is a conceptual language, not gibberish, every wff in it stands for (proxies) some construct. On the other hand $\mathscr{D}$ is not onto: there may be in C nameless constructs, such as the nonstandard objects in nonstandard analysis. Second, condition (v), sometimes called 'Leibniz' principle of the substitutivity of identicals', does not warrant any such substitutivety but rather the one of *different* expressions provided they happen to designate the same constructs. Third, this principle has nothing to do with reference. Hence it is not illustrated by ⌜The morning star = The evening star⌝. This last statement is, strictly speaking, false. What is true is that the two descriptions have the same referent – namely Venus. More on this in the next chapter. Fourth, the principle (v) has been challenged because it fails for the so-called "intensional contexts", i.e. in connection with "propositional attitudes" expressed by verbs like 'to know' and 'to believe'. For example, although '3' and 'III' designate the same number, the statement ⌜Everybody knows that 3 = III⌝ is plainly false (see, e.g., Linsky, 1967). We shall not take up this question since it belongs in pragmatics or in psychology (see Ch. 8, Sec. 4.3).

Fifth, the principle (v) is essential to all symbolic thinking. Its explicit statement should contribute to clarifying the way symbols perform their function, which is not to live their own life but to stand for, or deputize for, something else. Thus an equation such as ⌜$a = b$⌝ informs us that the letters 'a' and 'b', though different, name the same object. This trivial remark should help the beginning student of mathematics, who is often baffled by the apparent inconsistency of ⌜$a = b$⌝. And the remark should embarrass the nominalist, for in the formula '$a = b$' the two sides are not the same name. If mathematics were concerned with signs and languages rather than their designata, we should write: 'a' = 'b', which is false. In fact, as Frege (1879) noted long ago, we are here concerned with a single concept, now called 'a', now 'b'. (This does not endorse Frege's view that the theory of identity concerns only names or terms. The theory of identity is supposed to be universal, in the sense that it holds for objects of all kinds.)

Sixth and last, note that we have not defined the designation function $\mathscr{D}$: we have not specified the way it maps expressions into constructs. Any precise characterization of $\mathscr{D}$ involves the exact specification of such a correspondence, therefore of its domain Σ^{**} and range $\mathscr{P}(C)$ – hence a loss of generality. As soon as the designation function is specified, the conceptual language $\mathscr{L}_{KL}$ becomes a particular *semantic system.* A customary way of specifying ("defining") a semantic system is to lay down *designation rules* such as "Let 'M' designate the Newtonian mass function" (Carnap, 1942). But of course a formula such as this is a rule from a pragmatic point of view only, as it expresses a decision or convention. From a semantic point of view it is not a rule but a statement, namely an instance of "$\mathscr{D}(\sigma)=c$", where $\sigma\in\Sigma^{**}$ is a sign and c the construct σ designates. Besides, designation "rules", though necessary, are not sufficient for the purpose of formulating a body of factual knowledge, however modest it may be. Even an address book involves an additional semantic notion, namely that of *denotation*, a relation which goes from signs to factual items, either directly or via constructs. (More on denotation in Ch. 2, Sec. 2.4.) This relation includes the directory relation that matches names of people to names of places (addresses) and *represents* the physical relation between people and their dwellings:



In the case of scientific theories we have the composition of two relations: designation $\mathscr{D}$, from signs to constructs, and reference $\mathscr{R}$, from constructs to factual items. That is, in addition to the ordinary assumptions (construct-construct links) and the designation "rules", scientific theories also contain construct-fact correspondences. The latter tell us what the theories are about: which of its constructs refer to what and which represent what. In other words, the semantic systems of factual science, unlike those of pure mathematics, include semantic hypotheses or assumptions, often misnamed 'correspondence rules' (Chs. 2, 3, and 6). For this reason, and because a language should be a neutral instrument for expressing ideas, our Definition 1.4 of a conceptual language should not be mistaken for a definition of a factual theory. In other words, Definition 1.4 characterizes only a class of languages suitable for expressing a body of factual knowledge.

3. METAPHYSICAL CONCOMITANTS

3.1. *Basic Ontology*

Our definition of a language in Sec. 1.1 involved not only signs but also their denotata, i.e. the values of the coding function Δ. The denotata may be, but mostly are not, linguistic items. In fact the denotata of a sign system may be anything whatever: individuals, sets, relations, concrete or abstract, possible or impossible. They are *objects* in the general philosophical sense of the word, not in the sense of concrete tangible things.

We shall take this general notion of an object as primitive or undefined, for it is far too basic and important to be definable. And we may let ontology take care of its characterization even though we reject the very notion of a general theory of objects of any kind. In any case we shall assume the following partitions:

(i) Every *object* is either a factual item (e.g. an event) or a construct (e.g. a set) and none is both.

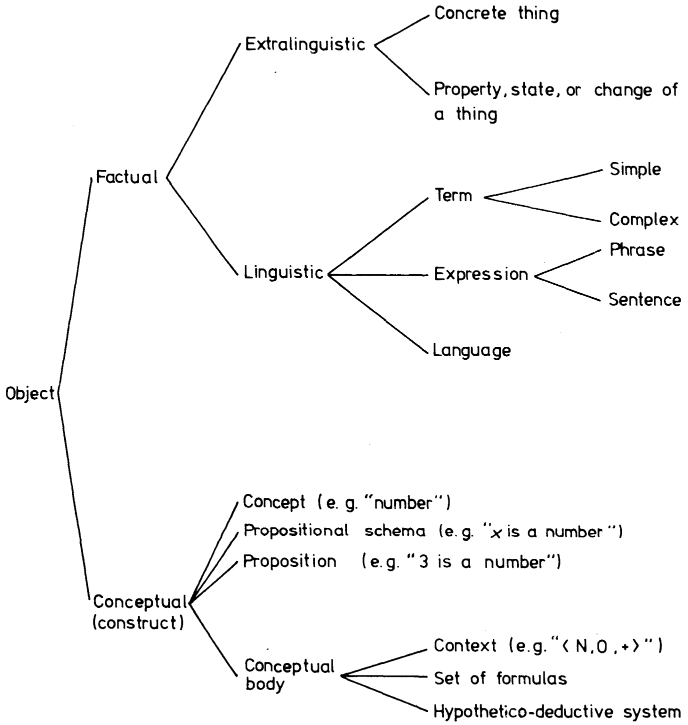
(ii) Every *factual object* is either linguistic or extralinguistic and none is both.

(iii) Every *linguistic object* is either a term (a member of some Σ or of some Σ^*) or an expression (an element of some Σ^{**}) or a whole language.

(iv) Every *construct* is either a predicate or a propositional function or a proposition or a set of either (with or without a structure).

In other words, the basic ontology accompanying our semantics admits the kinds of object shown overleaf.

That there is such a diversity of kinds of object is suggested by the fact that they satisfy radically different sets of laws: for example, the concepts of physics do not obey the same laws as their referents. Whether or not all objects in our ontology are attributed an *autonomous* existence is another matter – one for metaphysics not for semantics. However, the author hastens to declare that in his own metaphysics neither constructs nor even linguistic items are self-existing: the two are artifacts, hence dependent upon man – they were and are being created by mankind and will follow the fate of mankind. Surely signs, e.g. inscriptions, are physical objects: but they depend upon man for their coming into being as well as for their functioning as signs (proxies) for whatever they stand for. As to constructs, they are total fictions: what is real is the brain process that consists in thinking of some object. Let us elaborate on this point.


Object
Factual
Extralinguistic
Concrete thing
Property, state, or change of a thing
Linguistic
Term
Simple
Complex
Expression
Phrase
Sentence
Language
Conceptual (construct)
Concept (e. g. "number")
Propositional schema (e. g. "x is a number")
Proposition (e. g. "3 is a number")
Conceptual body
Context (e.g. "⟨ N, 0 , + ⟩ ")
Set of formulas
Hypothetico-deductive system

3.2. *Beyond Platonism and Nominalism*

Consider an external event such as a lightning bolt and a linguistic fact such as someone's uttering the sentence 'That was lightning'. Between these two physical processes we have an intermediate brain process – a thought. Change the speaker or the language and the sentence may change. However, we may surmise that, if the speakers' genetic make-ups and their funds of experience are similar, so will be their thought processes, whence their various utterances, though possibly different, will stand for the same statement or proposition. (See Fig. 1.1.)

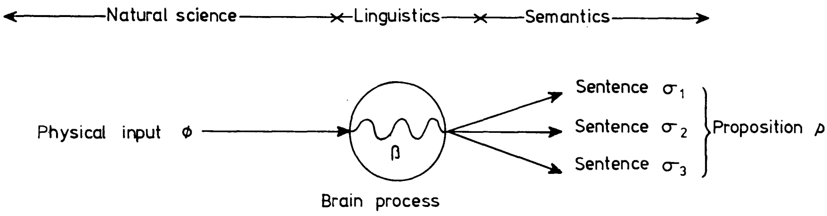


Fig. 1.1. A physical stimulus ϕ, the brain process β (thinking) it elicits, and its linguistic outputs σ_i, sentences expressing the proposition p.

Propositions are not physical objects: they have no reality aside from brain processes – just as there is no motion independent of moving things. It is a fiction, albeit an indispensable not an idle one, to assume that, in addition to the factual items and independently of them, there is such a thing as a proposition. For Platonists like Bolzano there are propositions *in themselves*, that need not have been thought by anyone, hence that may never be "discovered" (Bolzano, 1837). We make no such metaphysical assumption: we make instead the *methodological pretense*, or useful fiction, that things happen *as though* there were autonomously existing propositions (statements) as the designata of some sentences. (Cf. Sec. 1.2.) The concept of existence involved herein is that of conceptual existence not physical existence (Ch. 10, Sec. 4.1). We can turn the Platonist and the fictionalist loose in the field of constructs precisely because these *are* fictions. It is only where real (concrete, material) entities are concerned that the restrain advocated by the nominalist is in place.

Let us take another look at the situation depicted by Figure 1.1. The real events are the lightning, its perception, the thought process triggered in the subject by that perception, and his linguistic utterances. Each of these real events can be studied by some factual science. In particular, linguistics may take care of the sentences – and literary criticism may decree which sentences are "felicitous". Philosophical semantics starts thereafter, where linguistics leaves off. The former is not concerned with speech acts or even with linguistic items for their own sake or as constituents of human behavior but only insofar as they represent constructs.

So much so that philosophical semantics is hardly interested in nonconceptual signs such as 'ouch!' and 'liquid dream' or even in inscriptions, such as hieroglyphs. In other words, semantics is concerned with whatever grows at the tip of a conceptual sign. In this sense semantics is not a part of semiotics conceived as the science of signs (Morris, 1938). Philosophical semantics is a science of constructs and therefore may be regarded as a separate philosophical discipline or else as part of epistemology (see Ch. 10, Sec. 3).

Our view, though not Platonist, contrasts with the nominalist approach to semantics. Whether medieval or contemporary, materialist or empiricist, nominalism trusts the tangible (things and words) as much as it mistrusts the intangible – thoughts and constructs. Language is allegedly ostensible and controllable, whereas thought is hidden and can be wild, and constructs are admittedly fictions. The nominalist deals in terms and sentences not in concepts and propositions (see e.g. Zinov'ev, 1973). The motivations of nominalism are quite sound: to avoid obscurity, wildness, ghostliness, and Platonic hypostatizations. But the surgery it counsels, in particular the excision of constructs, eliminates the semantic problem instead of solving it. The "final solution", consisting in construing everything conceptual as a matter of signs, is in the same vein as the behaviorist proposal of regarding thinking as imperceptible laryngeal movements or as a disposition to talk: the former beheads semantics just as the latter decerebrates psychology. Whatever we may think of nominalist metaphysics, we cannot subscribe to nominalist semantics if we want to understand (*a*) that constructs are not entities on a par with things, (*b*) the very idea of a conceptual symbol as distinct from a nonconceptual one, (*c*) the dispensable character of every single symbol (but of course not of symbolization), (*d*) the fact that conceptual symbols must be adapted to the laws (logical, mathematical or philosophical) of constructs rather than the other way around, and (*e*) that these laws are not factual and are not subject to empirical tests.

The fruitful stand with respect to intangibles is not writing them off but to investigate whether they are lawful and whether positing them explains anything. Thus instead of outlawing the concept of a proposition, logicians have built propositional calculi. Likewise instead of ostracizing the concept of meaning for being intangible, we should clarify it by constructing a theory of meaning, just as scientists have been taming their own intangibles – fields, hydrogen bonds, gene pools, learning abilities, social structures, and the rest. Even linguists go beyond facts in order to understand them: in fact their mathematical models of language discard speakers and speech acts and they contain high brow constructs. Thus the set Σ^* of strings of a language (Sec. 1.1) is infinite; a phoneme is not a particular sound but an equivalence class of sounds – such as the m's uttered by different speakers; and the deep structure of a phrase is not perceptible either. A mathematical model of a language is not an empirical item but a more or less coarse representation of actual facts, to be tested by confrontation with empirical items. In the case of linguistics the empirical items are the finite samples of a language, such as those produced by Shakespeare, Johnny, and so on. Were the linguist to confine himself to such samples he would find hardly any stable and deep regularities: patterns, whether linguistic or chemical, must be hypothesized. Samples suggest and test a model but do not constitute it. Likewise if the semanticist were to limit himself to words and dictionaries he would never come up with theories of meaning and truth. His business is not to observe and describe speech acts but to analyze and systematize the semantic properties of concepts, propositions, and theories. By so doing he need not hypostatize constructs and meanings: conceptualists need not, nay they should not be Platonists. The conceptualist shares the nominalist's conviction that there are no universals in actual discourse – which is a string of concrete events (Goodman, 1951, p. 288). Only, he cannot see how universals could be dispensed with in *theorizing* about "word events", "illocutionary acts", or any other items.

We conclude with Table 1.2, which lists the main philosophies of the sign-construct relation to be discussed in this book.

TABLE 1.2

Sign and construct: main views

View	Sign	Concept, Proposition	Theory
Platonism	A physical object – the shadow of an idea.	A constituent of the Realm of Ideas, which is self-existent.	A body of ideas. Objects of a mathematical theory = Concepts.
Psychologism	A physical object acting as a proxy for some thought.	A thought of some kind.	A set of thoughts, actual or possible.
Nominalism (materialist or empiricist)	A physical (or alternatively experiential) item that stands in its own right or represents some other item of the same kind.	There is none.	A set of expressions: a part of a language. Objects of a mathematical theory = Signs (marks).
Conceptualist materialism	A physical object that stands for another object (or set of objects) – physical, mental, or abstract.	No constructs aside from mental objects, in turn brain processes. Construct = Equivalence class of brain processes – i.e. neither a concrete individual nor a Platonic Idea. For the sake of expediency pretend constructs exist by themselves.	A set of statements with deductive structure. Objects of mathematical theories = Constructs. Scientific theories have an external reference.

CHAPTER 2

REFERENCE

We propose to study now the semantical concept of reference – the *suppositio* of medieval logicians. This concept occurs in statements such as ⌜She was *referring* to witches⌝, ⌜Ecology *concerns* the organism-environment relationships⌝, ⌜Economics *deals with* the production and circulation of merchandises⌝, and ⌜Political science is *about* political institutions⌝. This semantic concept of reference should be distinguished from the psychological or pragmatic notion of reference involved in ⌜That theory suggests (or makes one think of) x⌝, ⌜This theory is meant to apply to x⌝, and ⌜The intended referent of x is y⌝. This other notion of reference is involved in learning how people actually create, learn, or use ideas, whereas the semantic concept of reference comes up when asking what a statement is about quite apart from the way it is conceived, applied, misapplied, or tested.

It might seem that an investigation of the semantic concept of reference should be trivial and therefore useless: don't we normally know what we are talking about? Unfortunately not: there is often endless argument as to what certain scientific theories are about. Hence we should welcome any semantic doctrine that could be of some help in spotting the genuine referents of a scientific theory. Since no such semantic doctrine seems to be available, we shall have to build one. But before proceeding to this task let a few examples vindicate our claim that the referents of a statement, or of a set of statements, are not always in evidence and that there are no obvious tools for bringing them to light.

1. MOTIVATION

Consider the following cases gleaned from the literature.

Case 1 Several philosophers have maintained that ⌜p is true⌝ is identical to ⌜p⌝, i.e. that the truth concept is redundant. On the other hand Bolzano held that the subject (referent) of ⌜p is true⌝ is p itself, which is not the case with the latter statement, whence the two are different. He even suggested the possibility of iterating this procedure, thus coming up with an infinite ladder of statements, every one of which has the previous one as its subject or referent (Bolzano, 1851, p. 85).

Case 2 The same philosopher had held previously that ⌜Some people are literate⌝ concerns only literate persons while ⌜Some people are not literate⌝ is about illiterates only (Bolzano, 1837, III, Sec. 305). True or false? And how about their equivalents ⌜Not everyone is illiterate⌝ and ⌜Not everyone is literate⌝ respectively?

Case 3 Aristotle taught that "A single science is one whose domain is a single genus" (*Post. Anal.*, Bk. I, Ch. 28). Right or wrong? How about ecology?

Case 4 Evolutionary biologists are divided on the question of the referents ("units") of population genetics and of the theory of evolution (see Williams, 1966, Ch. 4). Is it individual organisms, species, or populations? And does the theory state that selection acts on genotypes (individuals) or rather on phenotypes (populations)?

Case 5 Some proponents of the so-called identity theory argue that, although neurophysiology and psychology employ concepts with different senses, they have exactly the same referent, namely the person, whence those sciences constitute different ways of viewing the same facts, namely mental (= neurophysiological) events. This particular defense of the identity theory rests then on the semantic hypothesis that Referent = Fact. How about the ethology and the physiology of birds, which share their referents?

Case 6 Eminent physicists have claimed that the special theory of relativity is about the behavior of clocks and yardsticks. Others have held that it concerns observers in relative motion. Still others, that the referents of the theory are point masses inhabited by competent and well equipped experimenters communicating with one another through light signals. Finally, there are those who contend the theory to be about any systems connectible by electromagnetic signals. Take your pick.

Case 7 Most physicists act on the assumption that the quantum theory refers to autonomously existing microsystems such as neutrons and photons. But when it comes to "philosophizing" many claim that the theory is about sealed (unanalyzable) blocks constituted, in arbitrary proportions, by microsystems, measuring instruments, and observers.

Still others hold that the theory is concerned with the knowledge of nature rather than the latter (Heisenberg, 1958, p. 100).

Case 8 Consider the formula ⌜For all integers x: $x+1=1+x$⌝. Is it just about integers or rather about the whole system $\langle Z, 1, +\rangle$ (Rosenbloom, 1950, p. 110)? In any case, by which criterion is it either?

Case 9 The theory of control systems has been developed by engineers but has applications in biology and other fields of inquiry concerned with things, whether inanimate or alive, that have control devices built into them. Does the theory have a definite reference class at all?

Case 10 What is the reference class of elementary logic? Combinations of letters? Sentences? Propositions? Argumentative people? The world?

There should remain little doubt that identifying the referents of a statement or of a theory can be a thorny problem and that we need a semantic theory capable of helping us to perform that task. We shall presently expound one such theory. More precisely, we shall first study the rather unruly relation of reference and shall later on introduce a couple of law-abiding reference functions – one for predicates, the other for statements. The outcome may be regarded as a calculus of reference allowing one to compute the reference class of any composite statement as a function of the reference classes of its constituents. This calculus should help us to solve problems of referential ambiguity. By the same token it should free us from the recourse to authority as the "method" for ascertaining what scientific theories are about. On the other hand our calculus will not presume to tell us when a given predicate is applicable, i.e. what its correct reference or field of validity is: this is a business of science. In other words, we distinguish reference from extension – a subject to be studied in Ch. 9, Secs. 1 and 2.

2. THE REFERENCE RELATION

2.1. *An Unruly Relation*

We stipulate that the reference relation $\mathscr{R}$ holds between constructs (concepts, statements, or theories) on the one hand, and objects of any kind on the other. In other words, we adopt the following

CONVENTION The graph (or extension) of the reference relation $\mathscr{R}$ is a set of ordered pairs construct-object, i.e.,

$$\mathscr{E}(\mathscr{R}) \subseteq C \times \Omega, \quad \text{with} \quad C \subset \Omega,$$

where 'C' stands for the class of constructs and 'Ω' for the class of objects.

$\mathscr{R}$ has no simple formal properties. In particular $\mathscr{R}$ is not reflexive throughout its graph. For example, the concept of a star concerns stars rather than itself. On the other hand the number 7 refers to nothing. Nor is $\mathscr{R}$ either symmetric or antisymmetric. Finally $\mathscr{R}$ is not transitive either, as shown by the following counterexample:

$$r = \ulcorner\text{The statement below is false}\urcorner \tag{1}$$

$$s \tag{2}$$

where s refers to a third statement t. Clearly, although $\mathscr{R}rs$ and $\mathscr{R}st$, it is not the case that $\mathscr{R}rt$.

This last result has an important application in the foundations and philosophy of science, namely in relation to the semantic status of metanomological statements, or laws of laws (Bunge, 1961a; Angel, 1970). Consider the following propositions:

> A fundamental physical statement (e.g. equation) should involve no constants other than universal constants. (3)
>
> Newton's laws of motion are invariant under Galilei transformations. (4)
>
> Every quantum-mechanical formula should correspond to some classical formula. (5)
>
> If a local field theory is relativistically invariant, then it is also invariant under the combined conceptual inversion of charge, time, and parity. (6)

These statements and many others – some descriptive, some prescriptive – are often treated on a par with the object statements of a theory. However, it is apparent that they are *metastatements*, i.e. that they concern further statements. (Caution: not every metastatement belongs to some metatheory. None of the above does. Moreover some metastatements belong to object theories – as is the case with (4) and (6) above. For a typical confusion of 'metastatement' and 'metatheoretical statement' see Freudenthal (1971).) Further, since the reference relation is not transitive, the above propositions fail to refer to the same object their referents are about. Thus statement (4) is not a law of motion and (6) is not a field law – hence neither can be tested by observations on physical objects. Which suggests that, in spite of operationism, semantics should precede methodology: before we pose the problem of testing a statement we should know what it refers to. Enough of this for the moment.

In conclusion, $\mathscr{R}$ is neither reflexive nor symmetrical nor transitive – nor, indeed, seems to have any other definite formal trait. It is a wishy-washy relation and as such not amenable to theory. All our remarks in this section will therefore be intuitive. To obtain regularity we shall have to introduce a reference function. This will be done in Sec. 3; but before that we must take a closer look at $\mathscr{R}$.

2.2. *Immediate and Mediate Reference*

Consider a specific theory, or theoretical model t, of some concrete system s. Any such specific theory "defines" a model object m, or conceptual image of s, that is hoped to capture some of the traits of the latter. We may say that, while the *immediate referent* of t is the model object m, the ultimate or *mediate referent* of t is the real thing s (Bunge, 1967a). Table 2.1 exemplifies the idea.

TABLE 2.1

Examples of immediate and mediate reference

Theoretical model t or specific theory	Model object m = immediate referent of t.	Real system s = immediate referent of m. = mediate referent of t.
Enterprise theory	Directed graph	Enterprise
Contagion theory	Diffusion equation	Epidemics
Volterra's predation equations	Predator-prey system in constant environment	Rabbit-fox system or the like in wild life
Eye dioptrics theory	System of lenses	Mammalian eye
Magnetostatics of magnetic dipoles	Magnetic dipole	Terrestrial magnetic field

In every such case (*a*) the formulas of the theoretical model (specific theory) concern directly the model object itself, and mediately the thing modeled: for example, a theorem in enterprise theory may be about the degree of a vertex in the hierarchical tree of the enterprise; (b) the formulas are true of the model object (e.g. the unperturbed rabbit-fox economy) but only approximately true of the real thing – e.g. the rabbit-fox system in a variable environment subject to droughts, viruses, etc.; (*c*) the reference relation holds between every pair: $\mathscr{R}tm \,\&\, \mathscr{R}ms \,\&\, \mathscr{R}ts$ even though it is not transitive.

2.3. *Reference Class*

The set of referents of a given construct *c* is called its reference class. More explicitly, we introduce

DEFINITION 2.1 If *c* is a construct, then the *reference class* of *c* is the set of objects referred to by *c* (or the collection of items to which *c* bears $\mathscr{R}$), i.e. the relation class $[c]$ of *c* relative to $\mathscr{R}$:

If c is in C, then $[c] =_{\mathrm{df}} \{x \in \Omega \mid \mathscr{R}cx\} = \overleftarrow{\mathscr{R}}\text{‘}c$.

DEFINITION 2.2 A construct *c* *refers partially* to a class $A \subset \Omega$ iff A is included in the reference class of *c*, i.e., if $A \subseteq [c] = \overleftarrow{\mathscr{R}}\text{‘}c$.

Some constructs concern a single natural kind while others are about heterogeneous classes. (We do not apologize for employing the concept of a natural kind, essential to science, but we leave its elucidation to ontology.) For example, the reference class of "viscous" is the set of fluids, while that of "writing" is composed of the set of people and the set of written symbols. This difference is consecrated in the following

DEFINITION 2.3 A reference class is said to be *homogeneous* iff it is composed of elements of a single natural kind.

DEFINITION 2.4 A nonempty reference class is called *inhomogeneous* iff it is not homogeneous.

Every statement in factual science has a reference class assumed to be nonempty even though, on closer investigation, that class may prove void. For this reason it is often advisable to speak of *hypothetical* (or assumed) reference classes. A hypothetical reference class is often called an *intended* reference. It is not convenient to employ this second term in semantics, as it suggests the psychological concept of intention. (Whether or not somebody intends to entertain the hypothesis that a given construct has a certain reference class is a matter for psychology.) The class of objects for which a statement or a theory holds true may be called its *actual reference class*.

Since a scientific theory is a set of factual statements (with a deductive structure) every such theory has a reference class. This class is known or assumed to be nonempty. If the assumption has not yet been substantiated, then the theory itself may point the way to its referents. For example, a theory concerning a hypothetical extinct biological species will be instrumental in the search for fossil evidence relevant to the hypothesis. Should the hypothesis turn out to be false, the actual reference class of the theory would shrink to nought, but the assumed reference class would remain nonempty though indeterminate and idle, at least until further notice.

The reference class of a theory, whether it be the assumed or the actual one, is sometimes called the *ontology* of the theory. This is a misnomer, for an ontology is not a set of things but a philosophical theory concerning the basic traits of the world. Strictly speaking, the *ontology of a scientific theory* is the set of ontological (metaphysical, cosmological) hypotheses presupposed or countenanced by the theory. (See Bunge, 1973b.) For example, the reference class of classical electrodynamics is constituted by the set of bodies and the set of electromagnetic fields. On the other hand the ontology of the same theory is composed of wide ranging generalizations such as "Every thing is either body or field or a compositum of body and field", "The world is a plenum not a collection of atoms in void", "Except at boundaries all properties are continuous", "All properties are lawfully interrelated", and so on. Alternative systems of electrodynamics may or may not have different reference classes. If a change in scientific theory involves a modified reference class then the associated ontology may or may not change. (For example, action at a distance electrodynamics do not involve the field concept and consequently their ontology is closer to Greek atomism than to Aristotle's or Descartes' plenistic cosmologies. On the other hand the exact number of kinds of elementary particle and their precise properties are irrelevant to metaphysics.) One and the same scientific theory is consistent with a multitude of ontologies – all of a feather, though. Although science determines its own metaphysics, it does so only in outline: there is considerable leeway.

The preceding considerations on the relation between the reference class of a theory and its ontologies does not apply to formal science. The reference classes occurring in logic and mathematics are constituted by conceptual objects. For example, the reference class of the propositional calculus is the set of all propositions; number theory is concerned with numbers; topology is about topological spaces – and so on. In general: in formal science $\mathscr{R}$ never points to anything outside formal science. In other words, within formal science $\mathscr{R}$ pairs off constructs to constructs, i.e. $\mathscr{E}(\mathscr{R}) \subseteq C \times C$. To put it another way, we adopt

THESIS 2.1 *If c is a construct in formal science (logic or mathematics), then* $[c] = \overleftarrow{\mathscr{R}}`c \subseteq C$.

This is the postulate of the formal (nonfactual and a fortiori nonempirical) character of logic and mathematics. (Cf. Kraft, 1970.) If the thesis of the autonomy of formal science is accepted then it becomes clear (*a*) why the semantics of logic and mathematics, i.e. model theory (including Tarski's theory of truth), is hardly relevant to the semantics of factual science; (*b*) why logical and mathematical theories are not tested in the laboratory, and (*c*) why logic and mathematics have no associated ontologies – since ontologies are, by definition, theories about *entia* or existents. (If one insists on speaking of the 'ontology of formal objects' then one ought to identify it with formal science not with a branch of philosophy.) However, all this is controversial and will be taken up again, particularly in Ch. 10, Sec. 4.

2.4. *Factual Reference and Object Variable*

Factual items, such as things and events, are conveniently characterized as nonconceptual. To be more precise, we introduce

DEFINITION 2.5 The totality F of *factual objects* is the subset of Ω formed by nonconceptual elements and such that F is disjoint from C:

$$F =_{df} \{x \in \Omega \mid \neg(x \in C)\}.$$

If a concept, statement, or theory refers to one or more factual objects, it will be said to be *factual*. The subrelation $\mathscr{R}_F$ of factual reference pairs constructs off to factual objects, i.e. its extension is contained in the set of construct-fact pairs:

$$\mathscr{E}(\mathscr{R}_F) \subseteq C \times F.$$

This particular concept of reference comes with a particular concept of reference class as characterized by

DEFINITION 2.6 If c is a construct, then the *factual reference class* of c is the set of factual objects referred to by c:

$$c \in C \Rightarrow [c]_F = \overleftarrow{\mathcal{R}}_F{}^{\backprime} c =_{df} \{x \in F \mid \mathcal{R}cx\} \subseteq F.$$

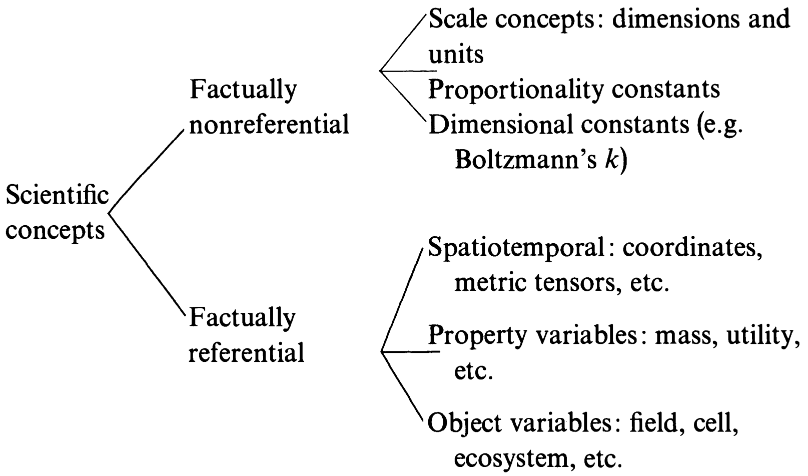
Factual science contains both factually referential concepts, like the one of attention, and concepts with an empty factual reference, such as "2". The same holds for metaphysics or ontology. Having a nonempty factual reference class suffices for a construct to belong either in science or in metaphysics. More explicitly, we state

THESIS 2.2 If $c \in C$ and $\emptyset \subset [c]_F \subseteq F$, *then* c *belongs in factual science or in metaphysics.*

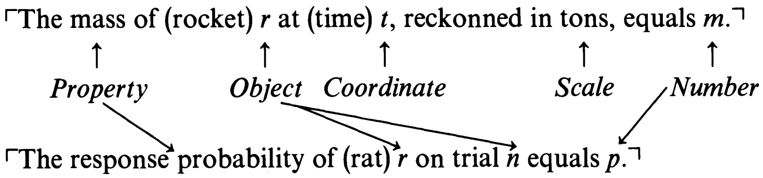
Having a precise mathematical form is no indication of ontological status: a factual concept, i.e. one with factual referents, may have a precise mathematical form while a nonfactual concept may fail to have one. Until one century ago some of the concepts of the infinitesimal calculus were in the latter condition.

Scientific concepts are often partitioned into constants and variables. Many of them, whether of the first kind or of the second, are quantitative concepts, and most of these are assigned dimensions and units. Thus the dimension value of a speed is LT^{-1}, and a standard speed unit is $cm\ sec^{-1}$. But dimensions and units give no precise indication of semantic status: any ratio of magnitudes with the same dimension value is dimensionless, hence unit free, even though it may concern some concrete object or other. And dimensional constants, like the gravitational constant γ, are not the value of a property of some physical system, so that they are devoid of factual reference even though they occur in law statements (Bunge, 1967b). A suitable semantic partition of the specific or technical concepts of factual science is rather the one shown overleaf.

What we have somewhat improperly called *scale concepts*, such as L and *cm*, have no factual reference. They are ingredients of some other concepts which do have a factual reference. Thus speeds, with dimension value LT^{-1} and reckoned on a suitable unit, are always speeds of something. This something, usually hinted at in the context, is of course the referent of "speed" – car, light wave, or what have you. In contrast, proportionality constants and dimensional constants, whether or not universal (independent of any particular material), are factually nonreferential: they concern nothing in particular. Thus unlike the velocity *of* light or the electric charge *of* the electron, Boltzmann's constant k is not the value of some property of a physical system.


Scientific concepts
Factually nonreferential
Scale concepts: dimensions and units
Proportionality constants
Dimensional constants (e.g. Boltzmann's k)
Factually referential
Spatiotemporal: coordinates, metric tensors, etc.
Property variables: mass, utility, etc.
Object variables: field, cell, ecosystem, etc.

The variables that are factually referential can in turn be classed into: (*a*) *object variables*, referring to things such as photons or persons; (*b*) *property variables*, representing properties of concrete things and relations among them, and (*c*) *spatiotemporal variables*, concerning the basic world framework. Variables of all three kinds are likely to occur in one and the same statement. Examples:


⌜The mass of (rocket) r at (time) t, reckonned in tons, equals m.⌝

⌜The response probability of (rat) r on trial n equals p.⌝
Property
Object
Coordinate
Scale
Number

The object variable (or constant, as the case may be) is not usually indicated in an explicit manner, except when a number of individuals (atoms, rats, nations) are involved, in which case each thing is assigned a numeral, a letter, or some other symbol – as is the case with ⌜P_i is the population of the *i*th country⌝. However, of all the variables the object variables are the most important. Without them the property variables would be baseless and the spatiotemporal variables would be purely mathematical constructs. Indeed, a property other than a purely formal (mathematical) property is a property of some concrete individual or other, whether actual or conjectural. And the spatiotemporal variables occurring in factual science have a similar though less apparent concrete support. Thus in physics every distance value is a distance between two points in some concrete system. And a time interval is the time separation between two events, one of which may be taken as the start of a process. No things, neither properties nor events – nor space or time. However, in the case of spatiotemporal variables it is usually possible and convenient to feign that the spatiotemporal framework is given prior to, and independent of, things and events. This pretense (the hypothesis of absolute space and time), if taken literally, ensues in an illusion of thinglessness, hence of absence of factual reference. The illusion is dispelled by a foundational analysis of spatiotemporal concepts and by the associated metaphysical construction of a relational theory of spacetime, i.e. one according to which space and time are defined on the class of facts. Without such a metaphysical groundwork the semantic analysis of statements containing spatiotemporal variables would be incomplete and misleading. Thus semantics and metaphysics, far from being mutually exclusive, are complementary. (More in Ch. 10, Sec. 4.)

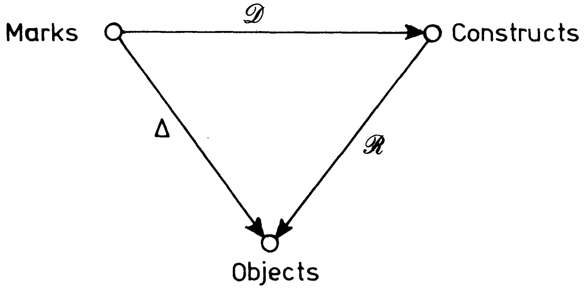
2.5. *Denotation*

Consider the following table.

Linguistic object—$\mathscr{D}$→	*Construct*—$\mathscr{R}$→	*Object*
Proper name	Individual concept	Individual
Predicate letter	Predicate	Property
Declarative sentence	Proposition	Set of facts

The relations $\mathscr{D}$ of designation and $\mathscr{R}$ of reference can be coupled to produce a relation that pairs off signs to objects. This relation, to be designated by Δ, will be construed as the (relational) product of $\mathscr{D}$ and $\mathscr{R}$. That is, we adopt

DEFINITION 2.7 Let Σ^{**} be the set of expressions of a conceptual language. The relation Δ, with domain Σ^{**} and codomain the set $\mathcal{P}(\Omega)$ of classes of objects, and such that $\Delta = \mathcal{D} \cap \mathcal{R}$, is called the *denotation* relation.



When the denotatum of a sign happens to be conceptual, Δ coincides with $\mathcal{D}$. Thus

'3' designates "3" = '3' denotes "3".

In other words, within formal science there is no difference between designation and denotation. But as soon as a construct refers to some extraconceptual object the difference becomes patent, and with it the impotence of the semantics of formal science to account for factual science. This concept of denotation has the advantage that it could not possibly be mistaken for the notions of reference and extension, as it so often happens with the presystematic notion of denotation (Geach, 1962). We can now use it although we shall not need it often.

2.6. *Reference and Evidence*

Reference classes should be kept distinct from bodies of evidence or evidence classes. Thus a hypothesis concerning the last glaciation period does not refer to the scars left by glaciers on rocks or to the distribution of certain plant species in our day. Likewise a theory about schizophrenia is not to be mistaken for a theory of the behavioral manifestations of it. This sounds obvious and has been pointed out before (Feigl, 1950), yet it has been controverted and is still being denied by die-hard operationists.

In general, the reference class of a scientific theory does not coincide with the body of evidence relevant to it – which may be nil. The two relations involved herein are radically different: whereas the relation of factual reference pairs off constructs to facts or things, the relation of evidence has the ***subset*** of observable facts as its domain and a subset of the totality of constructs as its codomain. (See Figure 2.1.) In other words, only some facts – namely those accessible to observation – count as evidence for or against a statement that refers to any given class of facts.

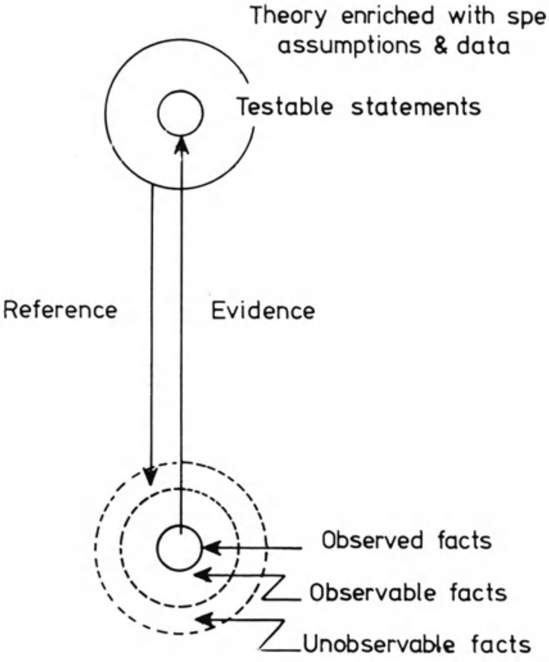

Theory enriched with specific assumptions & data
Testable statements
Reference
Evidence
Observed facts
Observable facts
Unobservable facts

Fig. 2.1. Reference is at best the converse of evidence.

And only some of the statements in a theory, even when adjoined specific assumptions and data, are susceptible to empirical test: all others must remain content with vicarious evidence, if any. (For details see Bunge, 1967a, Ch. 5 Sec. 5.6; Ch. 8 Sec. 8.4; Ch. 12; 1973a, Ch. 2; and 1973b, Ch. 10.) In sum *Reference ≠ Evidence*. Consequently *Reference class ≠ ≠ Evidence class*.

The differences between reference and evidence are best understood in relation to any theory concerning entities that are not accessible to direct observation, such as microphysical theories and economic theories. In this case the testability of the theory depends not so much on the size and kind of its reference class as on the existence of auxiliary theories capable of bridging the gap between the hypothetical referents and their observable concomitants, as well as on the experimental techniques capable of switching on and recording such unobservable-observable couplings.

The size of the reference class of a theory is irrelevant to the testability of the theory. Thus a theory explaining a unique process, such as the formation of our Moon, or the crumbling down of the American Empire, can have a sizable body of evidence. On the other hand a theory about an extremely large number of individuals, such as the neutrinos, is bound to have a comparatively poor evidence class due to the weakness of the interactions among neutrinos and other things. In short, the size of the reference class of a scientific theory furnishes no indication whatever about its testability. Conversely, the size of the evidence class of a theory is of little help in determining the reference class.

Consequently it is unnecessary, nay impossible, to elaborate on the testability-reference relation. All we can do is to issue warning notices and to enunciate the following

THESIS 2.3 If a theory is empirically testable then it has a nonempty hypothetical reference class.

The converse is false: hypothesized reference does not ensure testability. (Thus we can imagine, and theorize on, things that are in principle inscrutable, like the soul of plants.) Nonempty reference is then necessary for empirical testability. But it is insufficient. Consequently before we inquire into the possibility of testing a theory we must find out what it refers to. (Imagine trying to design an experiment for testing a hypothesis with no definite reference, such as a conjecture about a possible world.) Consequence for philosophy: *The semantics of science should precede the methodology of science.* This conclusion contradicts the so-called verifiability doctrine of meaning, according to which the meaning (and in particular the reference) of a statement consists in the way it is verified or, more generally, put to the test. This doctrine has constituted a stumbling block in the development of semantics, for it has mixed up questions of semantics, such as those of sense and reference, with problems of methodology, such as that of empirical testability. (Worse: it has left the latter unanalyzed.) Positivist semantics has had its chance and has failed. Give now realist semantics a chance.

2.7. *Misleading Cues in the Search for Referents*

Before taking the trouble to build a theory of reference we should check whether any of the existing views on the matter might help us to spot the referents of any given scientific formula. Let us therefore make a quick – hence necessarily somewhat unfair – review of the most popular opinions on the subject.

(i) *Grammaticalism:* "The denotatum of a sentence is always denoted by the grammatical subject". This view is false if grammatical predicates are regarded as unary, partially true if their actual complexity is taken into account. Thus traditional grammar would tell us that the referent of ⌜Woman is sweeter than man⌝ is the fair sex, while in point of fact the statement refers to both sexes. A suitable enlargement of the notions of grammatical predicate and subject might help us out in the case of ordinary language expressions. But it would be insufficient for spotting the referents of a statement clad in mathematical language: grammar cannot and does not want to attain the fine structure accessible to mathematics. It would be pointless to ask the grammarian to find the reference class of, say, a Gaussian distribution or a differential equation. Only an analysis of the whole theory in which the formula occurs could produce the desired result. And such an analysis, which lies within the scope of the semantics of science, is beyond the scope of grammar – even of the Chomsky type.

(ii) *Psychologism:* "Search for the intentional object, i.e. for the object of consciousness associated with the given construct". This recipe of Brentano's is definitely misleading: one and the same formula can be "read" in different ways by different persons. In other words, the intended referent of a construct depends not only on the construct but also on the person who thinks of it and the circumstances under which he or she is thinking. This is of interest to psychology not to semantics: the latter is concerned with the hypothetical though possibly real referents of constructs.

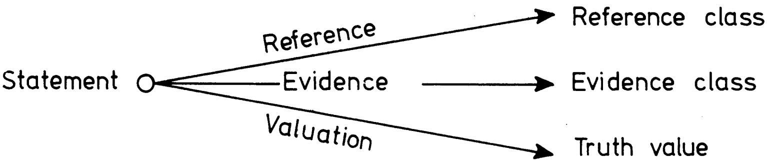
(iii) *Linguistic pragmatism:* "No expression refers by itself: it is only some user who, under definite circumstances, can attach a referent to an expression. Hence the act of referring, not reference, should be the object of semantic analysis". While it is true that expressions have no reference unless someone uses them, there is nothing wrong with abstracting from users and circumstances to speak of the reference class of a construct as a fixed set. Surely scientific theories would not exist unless there existed beings capable of making and using them. But the objectivity of such theories depends on their being assigned reference classes assumed (rightly or wrongly) to consist of real entities. (The concept of semantic objectivity is involved herein. That of methodological objectivity will concern us in Ch. 10, Sec. 1.2.) Paradoxically enough, there is no objectivity without abstraction from users and circumstances. If someone claims that a certain theory concerns *A*'s rather than *B*'s, he is expected to substantiate this claim by analyzing the initial assumptions of the theory rather than his own speech acts. Reference hypotheses must be impersonal and open to public inspection.

(iv) *Classical operationism:* "The meaning (=denotation) of a piece of science is to be sought in the operations aimed at testing it: a formula is meaningless unless the procedures for verifying it are specified." This thesis boils down to the conflation of reference and evidence, criticized in Sec. 2.6. Besides, the operationist criterion does not help us in spotting the hypothetical referents of a predicate, for predicates are not testable. And it is misleading in relation to statements, for it diverts our attention from reference to evidence: from the real thing (e.g. loving) to its observable manifestations (e.g. blushings), from facts to documents about them. Empirical tests must be preceded by reference analyses: without knowing or assuming what a given statement is about we would not know how to put it to the test.

(v) *Frege:* "Statements denote (*bedeuten*) truth values". While names and definite descriptions can denote objects, the denotation (*Bedeutung*) of a statement is either truth or falsity. Thus whereas the description "2^2" denotes the number four, the statement $\ulcorner 2^2=4 \urcorner$ denotes truth, and so does $\ulcorner 2>1 \urcorner$. Consequently

$$(2^2=4)=(2>1)$$

"is a correct equation" (Frege (1891) in Angelelli, p. 132). The ambiguous use of the term 'denotation' and the absurdity it leads to would deserve no mention were it not that this slip has been blown up into a doctrine that has misled many a contemporary semanticist. Moral: Keep reference and truth value separate. (Another: Honor the great man's hits not his misses.) A statement *has* a reference class, may be *supported* (or undermined) by an evidence class, and it may be *assigned* a truth value.



(vi) *Phenomenology:* "Referents are a nuisance: just bracket them out". According to Husserl, to get a direct spiritual vision of an essence (*Wesensschau*) we must discard (*ausklammern*) the external world and drop all previous knowledge of the object. No comment.

Conclusion: The popular views on reference, in particular operationism and Wittgenstein's, do not help us to identify the referents of a construct. Moreover they are misleading. Worse: neither do any of the elaborate and articulate systems of semantics help us out, particularly in relation to scientific constructs. Thus Carnap (1942, 1947), whose work was no doubt exact, elaborate and systematic, did not propose a theory of reference and was hardly interested in our subject. This may have been due to the mistaken impression that reference presents no real problem. Which may be true in the cases of ordinary knowledge and mathematics. But scientific constructs, which dwell between those two covers, do pose a challenge, as we saw in Sec. 1. Let us accept this challenge.

3. THE REFERENCE FUNCTIONS

3.1. *Desiderata*

Consider the table on the next page.

A first point to note is that we are tacitly distinguishing reference from extension. Thus in the 2nd row we state that "Clairvoyance" concerns mind readers and the like without claiming that there are any such

	Construct	Referent(s)
1	Human	Mankind
2	Clairvoyance	Mind readers, soothsayers, etc.
3	Conducting	Bodies
4	The moon is round	Moon
5	The moon is round or the cat is fat.	Moon and cat
6	The moon is round and the cat is fat.	Moon and cat
7	All quarks are charged.	Quarks
8	Nobody has so far observed any quark.	Quarks and people
9	Some people loathe animals.	People and animals
10	Nobody loathes animals.	People and animals

individuals. Moreover, if pressed we might grant that the extension of that predicate is nought, while insisting that its assumed reference class is composed of the individuals who claim to have that ability. Likewise in the 7th row we talk about the hypothetical entities called quarks, which at the time of writing have not been identified in the laboratory: the extension of "Quark" is not known. In short, as a rule $\mathscr{R}(c) \neq \mathscr{E}(c)$, where $\mathscr{R}$ and $\mathscr{E}$ designate the reference and the extension functions respectively. In this we follow Buridan (cf. Geach, 1962).

Secondly, it can be observed that reference is rather insensitive to the propositional connectives, i.e. to the gross propositional structure of a statement. Look at lines 5 and 6: the disjunction and the conjunction of two given statements are assigned the same referents. And lines 9 and 10 suggest that a statement and its negate should be assigned the same referents – which is reasonable since they are about the same things. Finally, lines 7 and 8 suggest that a theoretical statement should not be mistaken for a pragmatic one: while the former is supposed to have an objective reference, the latter is bound to concern, at least partly, the cognitive subject.

The above example and others suggest that a theory of reference should satisfy the following desiderata.

*D*1 The theory should define at least one reference function, not just a relation, if it is to enable one to compute the reference class of a complex predicate or formula out of the assumed reference classes of its components.

*D*2 The range of the reference function(s) should be a set of sets rather than the collection Ω of objects, since a single construct may concern a whole class of things.

*D*3 The reference class of a statement and of its negate should be the same.

*D*4 The reference class of an arbitrary propositional compound should equal the union of the reference classes of the components.

*D*5 The reference class of a relation should equal the field of the latter.

The functions to be introduced in the next subsection will prove to satisfy the preceding desiderata.

3.2. *Principles and Definitions*

We need two reference functions, one for predicates, the other for statements. These two types of construct are related in the following way (cf. Ch. 1, Sec. 2.2). An n-ary predicate P is a function

$$P: A_1 \times A_2 \times \cdots \times A_n \to S$$

from n-tuples of objects (individuals) to statements, such that the value of P at $\langle a_1, a_2, \ldots, a_n \rangle$ in $A_1 \times A_2 \times \cdots \times A_n$ is the atomic statement $Pa_1 a_2 \ldots a_n$ in S. The reference function for predicates will be defined on a set of predicates, and the one for statements will be defined on the set of statements formed with those predicates. The former is introduced by

DEFINITION 2.8 The *reference class of a predicate* is the collection of its arguments. More precisely, let $\mathbb{P}$ be a family of n-ary predicates with domain $A_1 \times A_2 \times \cdots \times A_n$. The function

$$\mathcal{R}_p : \mathbb{P} \to \mathcal{P}\Big(\bigcup_{1 \leqslant i \leqslant n} A_i \Big)$$

from predicates to the power set of the union of the cartesian factors of the domains of the former is called the *predicate reference function* iff it is defined for every P in $\mathbb{P}$ and

$$\mathcal{R}_p(P) = \bigcup_{1 \leqslant i \leqslant n} A_i .$$

Example 1 $\mathcal{R}$(Mutating) = Organisms. *Example 2.* Define a binary relation $<$ on some set A. Then the reference class of $<$ is A. *Example 3*. Let $F: A \rightarrow B$ be a function. Since F is a particular case of a relation on $A \times B$, $\mathscr{R}_p(F) = A \cup B$.

DEFINITION 2.9 Let $\mathbb{P}$ be a family of n-ary predicates with domain $A_1 \times A_2 \times \cdots \times A_n$ and let S be the totality of statements formed with them. The function

$$\mathscr{R}_s : S \rightarrow \mathscr{P}\Big(\bigcup_{1 \leqslant i \leqslant n} A_i \Big)$$

is called the *statement reference function* iff it is defined for every s in S and satisfies the following conditions:

(i) The referents of an *atomic statement* are the arguments of the predicate concerned. More precisely, for every atomic formula $Pa_1 a_2 \ldots a_n$ in S

$$\mathscr{R}_s(Pa_1 a_2 \ldots a_n) = \{a_1, a_2, \ldots, a_n\}.$$

(ii) The reference class of an arbitrary *propositional compound* equals the union of the reference classes of its components. More exactly, if $s_1, s_2, \ldots, s_m$ are statements in S and if ω is an m-ary propositional operation,

$$\mathscr{R}_s(\omega(s_1, s_2, \ldots, s_m)) = \bigcup_{1 \leqslant j \leqslant m} \mathscr{R}_s(s_j).$$

(iii) The reference class of a *quantified formula* equals the reference class of the predicate occurring in the formula. More explicitly, if P is an n-ary predicate in $\mathbb{P}$, and the Q_i, for $1 \leqslant i \leqslant n$, are arbitrary quantifiers, then

$$\mathscr{R}_s((Q_1 x_1)(Q_2 x_2) \ldots (Q_n x_n) Px_1 x_2 \ldots x_n) = \mathscr{R}_p(P).$$

Example 1. $\mathscr{R}_s$(Venus is larger than Mars) = {Venus, Mars}.

Example 2. $\mathscr{R}_s(p \,\&\, q) = \mathscr{R}_s(p \vee q) = \mathscr{R}_s(p) \cup \mathscr{R}_s(q)$.

Example 3. $\mathscr{R}_s$(All humans are mammals) = Animal Kingdom.
More on this in the next subsection and in Sec. 5.2.

The relations among the various sets and functions involved in the preceding definitions are summarized in the diagram on the next page.

$$\begin{tikzcd}
\text{Predicate } P \arrow[r, "\mathcal{R}_p"] & \bigcup\limits_{1\leqslant i\leqslant n} A_i \\
\text{Domain } A_1\times A_2\times\cdots\times A_n \arrow[r] & \text{Statement family } S \arrow[u, "\mathcal{R}_s"']
\end{tikzcd}$$

Remark 1 Constructs other than predicates and statements, such as the individual x and the set y occurring in the statement $\ulcorner x\in y\urcorner$, are assigned no referents in our theory. The reason is that individuals and collections thereof are to function as referents themselves. *Remark 2* According to Definition 9i, $\ulcorner Pa\urcorner$ is taken to concern the object named a, not the name 'a'. One might be tempted to introduce a new symbol, say 'a', for the nominatum of a. But this would start an infinite regress. *Remark 3* Our analysis of reference is applicable in all contexts, whether or not "extensional" (truth functional). For example,

$$\mathcal{R}(\text{Smith knows that } p)=\{\text{Smith}, p\}.$$

A binary epistemic predicate E, such as "knowing", "believing", or "doubting", will be analyzed as a function

$$E\colon \text{People}\times\text{Statements}\rightarrow\text{Statements},$$

whence

$$\mathcal{R}(E)=\text{People}\cup\text{Statements}.$$

There is nothing opaque about "oblique" reference if analyzed as multiple rather than single. More on "intensional" contexts in Ch. 4, Sec. 1.3. *Remark 4* Our definitions of the reference functions are necessary but not sufficient for a correct identification of the referents of a construct. A precise identification requires some fund of knowledge and it depends upon one's metaphysics. Consider, indeed, the statement that the weather is fine. The *ostensive* (or manifest or superficial) referent of this statement is the weather. But the weather is not a thing: it is a state of a thing. The *deep* (or hidden or genuine) referents of the given statement are the terrestrial atmosphere and mankind. In fact the given proposition may be construed as short for ⌜The atmosphere is in a state that suits humans⌝. The metaphysical principle behind this analysis of the given proposition is that the referents of factual constructs are concrete systems (things). For a radically different view concerning the relations of semantics to metaphysics, see Shwayder (1961).

3.3. *Some Consequences*

We proceed to drawing a few consequences from our Definitions 8 and 9 of the reference functions for predicates and for statements.

COROLLARY 2.1a The reference class of the equality (inequality) relation equals the set A on which $=$ is defined:

$$\mathcal{R}_p(=)=\mathcal{R}_p(\neq)=A.$$

COROLLARY 2.1b The reference class of a statement of inequality (equality) is constituted by the individuals concerned. That is, if $a, b \in A$, then

$$\mathcal{R}_s(a\neq b)=\{a, b\}, \qquad \mathcal{R}_s(a=b)=\{a\}.$$

COROLLARY 2.2a The reference class of the membership relation $\in$ equals the universe U of sets on which it is defined:

$$\mathcal{R}_p(\in)=U.$$

COROLLARY 2.2b The reference class of a statement of class membership is constituted by the relata concerned:

$$R_s(a\in A)=\{a, A\}$$

Remark. This last corollary gives rise to an apparent paradox. A membership statement is equivalent to a predicative statement:

$$Pa \quad \text{iff} \quad a\in\mathcal{E}(P), \quad \text{where} \quad \mathcal{E}(P)=\{x \mid Px\}$$

is the extension of the predicate. But $\mathcal{R}_s(Pa)=\{a\}\neq\mathcal{R}_s\,(a\in A)$. That is, *equivalents may have different reference classes*. And why should they unless equivalence be mistaken for identity or reference class for extension? ⌜Pa⌝ is equivalent to but not identical with ⌜$a\in\mathcal{E}(P)$⌝, if only because the predicates involved in each statement are not only of different rank and different type: while P applies to individuals, $\in$ relates individuals to sets and thus mixes objects of different types. Moral: The reference class we assign to a statement depends on the way we analyze the latter.

COROLLARY 2.3a The reference class of a set relation equals the union of the sets involved: If ρ is a binary relation between the sets A and B, then

$$\mathcal{R}_p(\rho)=A\cup B.$$

COROLLARY 2.3b The reference class of a statement of class relation consists of the classes involved:

$$\mathcal{R}_s(\rho AB)=\{A,B\}.$$

In particular

$$\mathcal{R}_s(A\subseteq B)=\{A, B\}$$

COROLLARY 2.4a The reference class of the conditional predicate $P\Rightarrow Q=\neg P\vee Q$ equals the domain shared by P and Q. More explicitly: if $P{:}A\rightarrow S$ and $Q{:}B\rightarrow S$ are unary predicates in $\mathbb{P}$, then because $P\Rightarrow Q{:}A\cap B\rightarrow S$,

$$\mathcal{R}_p(P\Rightarrow Q)=\mathcal{R}_p(P)\cap\mathcal{R}_p(Q)=A\cap B.$$

COROLLARY 2.4b The reference class of a universal conditional equals the intersection of the reference classes of the predicates involved:

$$\mathcal{R}_s((x)\,(Px\Rightarrow Qx)=\mathcal{R}_p(P)\cap\mathcal{R}_p(Q)).$$

Example The reference class of ⌜All ravens are black⌝, as well as of its negate, is the class of birds not that of black things nor any other one. Indeed, the predicate "$R\Rightarrow B$" is defined on the intersection of the domains of R and B, i.e. on $\mathcal{R}_p(R)$, because the domain of R is Birds and that of B is Things. More in Sec. 5.2.

COROLLARY 2.5a The reference class of a tautological predicate equals the union of the reference classes of the component predicates, whence it is not generally empty. In particular

(i) $\mathcal{R}_p(P\vee\neg P)=\mathcal{R}_p(P)$,
(ii) $\mathcal{R}_s((P)_{P\in\mathbb{P}}(P\vee\neg P))=\bigcup_{P\in\mathbb{P}}\mathcal{R}_p(P)$.

COROLLARY 2.5b The reference class of a tautological statement equals the union of the reference classes of the predicates involved, whence it is not generally empty. In particular

(i) $\mathcal{R}_s((x)(Px \vee \neg Px)) = \mathcal{R}_p(P)$,
(ii) $\mathcal{R}_s((P)_{P \in \mathbb{P}}\,(x)(Px \vee \neg Px) = \bigcup_{P \in \mathbb{P}} \mathcal{R}_p(P)$.

Remark 1 The tautology in (i) is *valid in the domain* of P, while the one in (ii) is valid in the total domain of the predicate family $\mathbb{P}$, i.e. *valid* without further ado. *Remark 2* Reference class and extension coincide only in the case of tautological predicates, or of tautologies, involving unary predicates. This coincidence seems to be an important source of the common confusion between reference class and extension.

COROLLARY 2.6a The reference class of a predicate and its negate are the same: If P is in $\mathbb{P}$, then

$$\mathcal{R}_p(P) = \mathcal{R}_p(\neg P).$$

COROLLARY 2.6b The reference class of a statement and of its denial are the same: if p is in S, then

$$\mathcal{R}_s(p) = \mathcal{R}_s(\neg p).$$

Example $\mathcal{R}_s$ (Some people are cruel) $= \mathcal{R}_s$ (Nobody is cruel) = People.

COROLLARY 2.7 The reference class of a singular tautological statement, or of its negate (the corresponding contradiction), equals the union of the reference classes of its component. In particular

$$\mathcal{R}_s(p \vee \neg p) = \mathcal{R}_s(p \mathbin{\&} \neg p) = \mathcal{R}_s(p) = \mathcal{R}_s(\neg p).$$

Remark 1 It is often claimed that analytic statements do not inform about the world because they are not about it. While the claim is right the reason given in support of it is dubious. ⌜Antarctica is cold or it is not cold⌝ concerns Antarctica, both for common sense and on our semantics, even though it "says" nothing, i.e. its factual content is nil. It is for the latter reason, not because of its alleged lack of reference, that the tautology is not affected by the recent finding that Antarctica was not cold during the Triassic. The same holds for analytically false statements: if singular they refer even if they do not inform. *Remark 2* Likewise for tautological predicates: since they hold for anything they refer to everything. *Remark 3* The preceding consideration would seem to contradict the thesis of the autonomy of logic espoused in Sec. 2.3. It does not, for what is peculiar to tautological singular statements is their *validity irrespective of their reference*, hence the impossibility of checking them against experimental evidence. *Remark 4* Since any given tautology is equivalent to every other tautology, it would seem that Corollary 7 leads to contradiction. It does not because we have refrained from requiring that equivalents be coreferential. Recall the comment to Corollary 2b. *Remark 5* The preceding remarks show the need for a concept of strong equivalence, or equivalence both syntactical and semantical. A possible definition is this: Two statements are said to be *strongly equivalent* iff they are equivalent and have the same referents. However, we shall not pursue this matter here. *Remark 6* Our reference functions refer to (are defined on) predicates and statements not on signs such as predicate letters and sentences. While on Definitions 8 and 9 every predicate or statement is referential (even if it concerns the null individual), not every expression denotes. A nonsense sign such as '&(%+' presumably occurs in no conceptual language, hence it designates nothing, whence it denotes nothing. (But of course it may suggest something – e.g. that the linotype ran amok – and may thus have a pragmatic meaning.) Even sophisticated signs composed of individually significant marks in a way that looks well formed may fail to designate constructs. (Two famous frauds are '1/0' and '$\{x \mid \neg(x \in x)\}$'. These are not, as is sometimes held, irreferential concepts: they are just nonconceptual signs.) *Remark 7* Even logical predicates refer – namely to constructs. For example, the reference class of the conjunction is the set of all the statements on which it is defined. This follows from Definition 8 applied to & construed as a certain function $\&: S \times S \rightarrow S$.

What follows will not depend critically on the precise formulation of our definitions of $\mathscr{R}_p$ and $\mathscr{R}_s$: in most cases it will suffice to assume that these functions are somehow well defined. Moreover, in the sequel we shall hardly distinguish between these two functions, for we shall be concerned with constructs in general.

3.4. *Context and Coreference*

The concepts of reference functions and reference class allow us to elucidate a number of other concepts. For example the following notions.

DEFINITION 2.10 The ordered triple $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ is called a (conceptual) *context* (or *frame*) iff S is a set of statements in which only the predicate constants in the predicate family $\mathbb{P}$ occur, and the reference class of every P in $\mathbb{P}$ is included in the universe or domain $D \subseteq \Omega$.

DEFINITION 2.11 c is called an *individual construct* in the context of frame $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ iff (i) c is either in $\mathbb{P}$ or in S, and (ii) the reference class of c is a unit set or singleton.

DEFINITION 2.12 c is called a *universal construct* in the context or frame $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ iff c is in $\mathbb{P}$ or in S but is not an individual construct.

Remark The preceding definitions of individual and universal relativize these concepts to a definite conceptual context or frame. This is as it should be, for one and the same object may be regarded as an individual in a given context and as a set in another. And a property may concern every individual in a class but not all individuals. This gives some precision to the classical distinction formulated by the Philosopher: "By the term 'universal' I mean that which is of such a nature as to be predicated of many subjects, by 'individual' that which is not thus predicated. Thus 'man' is a universal, 'Callias' an individual" (*On Interpretation*, Ch. 7).

DEFINITION 2.13 c is said to be an *individual construct* iff c is an individual construct in every context.

DEFINITION 2.14 c is said to be a *universal construct* iff c is a universal construct in every context.

Whether these last two concepts have any application must be left to other disciplines to find out. (Perhaps their only use is to occur in the conjecture that there are no such constructs.) Let us now turn to the comparison of reference classes.

DEFINITION 2.15 Two constructs are called *coreferential* (or *equirefetial*) in a context $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ iff (i) they belong to the context $\mathbb{C}$ and (ii) they have the same reference class. Symbolically:

$$\text{If } c \text{ and } c' \text{ are in } \mathbb{C}, \text{ then } c \sim_r c' =_{df} \mathcal{R}(c)=\mathcal{R}(c').$$

Remark The relativization to a context is not really needed if one proceeds carefully – i.e. with hindsight. For example, in the context of nonrelativistic physics the concepts of mass and charge are equireferential: both concern only bodies. But in relativistic physics they cease to be coreferential (although they are still referentially commensurable). For, although the two concepts concern bodies, that of mass concerns now also reference frames, and it is possible to make certain fields serve as frames. Of course two different concepts of mass are involved in the two contexts even though the two are usually designated by the same word. Should this be overlooked, the relativization to a context would avoid further confusion. (More in Sec. 4.2.)

As is clear from Definition 15, the relation $\sim_r$ is an equivalence relation. As such it enables us to form equivalence classes. Thus it allows us to form

DEFINITION 2.16 The set $[c]_{\mathbb{C}}$ of *coreferentials* of a given construct c in a context $\mathbb{C}$ is

$$[c]_{\mathbb{C}} =_{df} \{c' \in \mathbb{C} \mid c' \sim_r c\}.$$

Example All of the theories of solids, no matter how different, have the same reference class, namely the set of solid bodies.

Being an equivalence relation, equireference induces a partition of the totality C of constructs, i.e. an exhaustive division of C into mutually disjoint classes of coreferentials. This partition or collection of equivalence classes is the quotient set of C by $\sim_r$, or $C/\sim_r$ for short. We may call $\Pi = C/\sim_r$ the *referential partition* of the conceptual universe. This concept allows us to elucidate the intuitive concept of semantic homogeneity of a theory (Bunge, 1967a, Ch. 7) by means of

DEFINITION 2.17 A theory will be said to be *referentially homogeneous* iff it is included in an element of the referential partition $C/\sim_r$, i.e. if its constructs are all in one of the equivalence classes of equireferentials.

DEFINITION 2.18 A theory will be said to be *referentially heterogeneous* iff it is not referentially homogeneous.

Example Psychology and physics are referentially heterogeneous. On the other hand, according to the identity theory, psychology and

neurology are referentially homogeneous since both concern animals.
Let us next specialize the reference functions.

4. FACTUAL REFERENCE

4.1. *The Factual Reference Class*

In the so-called empirical sciences we are particularly interested in a special concept of reference, namely reference to concrete (though perhaps hypothetical) existents. We therefore introduce

DEFINITION 2.19 Let $\mathscr{R}(c)=A_1\cup A_2\cup\ldots\cup A_f\ldots\cup A_n$ be the reference class of a construct c, where the A_i for i between 1 and $f<n$ are sets of non-conceptual items, i.e. such that $A_i\not\subset C$. Then the *factual reference class* of c is the union of the classes of factual items:

$$\mathscr{R}_F(c)=A_1\cup A_2\cup\ldots\cup A_f\subseteq\mathscr{R}(c).$$

Example 1 $\mathscr{R}_F$ (Light rays are represented by straight lines) = Light rays. The total reference class includes also the set of straight lines. *Example 2* Let $f: A\rightarrow B$ be a function from a set A of concrete systems to a set B of numbers. For example, f could represent the relative dilation of bodies. Then $\mathscr{R}_F\ (f)=A$, whereas $\mathscr{R}_p\ (f)=A\cup B$. *Example 3* Let $g: A\times B\rightarrow R$, where A is a set of concrete systems, B the set of conceivable scale-cum-unit systems associated to the magnitude g, and R the set of real numbers. Since B is conventional, it is included in the class of constructs, so that we are left with $\mathscr{R}_F(g)=A$.

Not all the constructs occurring in factual science have a factual reference. For one thing the logical concepts, such as "not" and "all", have no such reference. Nor do scale concepts, proportionality constants, and dimensional constants have referents (recall Sec. 2.4). It will therefore be convenient to coin a name for such constructs. The following conventions will do.

DEFINITION 2.20 A construct which is either tautologous or has an empty factual reference class is said to be *factually vacuous*.

DEFINITION 2.21 A construct which is not factually vacuous, i.e. which is neither tautologous nor has a void factual reference class, is called *factual*.

Example 1 ⌜1 m = 100 cm⌝ is factually vacuous. *Example 2* The state function ψ is a factual construct because it concerns physical entities. *Example 3* The semantic assumption ⌜ψ represents the state of an individual microsystem⌝ is a factual construct because it is nontautologous and includes the factual concept ψ.

Let us now restrict the concept of a context introduced by Definition 10:

DEFINITION 2.22 The triple $\mathbb{C}_F = \langle S, \mathbb{P}, D \rangle$ is called a *factual context* iff (i) it is a context and (ii) $\mathbb{P}$ includes a nonempty subset of factual predicates.

DEFINITION 2.23 Any context that fails to be factual is called a *formal context*.

DEFINITION 2.24 Let $\mathbb{C}_F = \langle S, \mathbb{P}, D \rangle$ be a factual context. The predicate in $\mathbb{P}$ with the largest factual reference class is called *maximal* in $\mathbb{C}_F$. Symbol: $P_{\max}$.

Remark 1 Logic, mathematics and semantics are formal contexts since they deal with constructs. On the other hand history, linguistics, epistemology and metaphysics are factual contexts. Mythology, religion and science fiction are factual contexts too. Their extension or domain of truth is another matter. *Remark 2* Definition 23 effects a dichotomy in the set of contexts. However, the factual/formal dichotomy is not universally accepted: to some even mathematics is *au fond* factual or empirical, while a few hold science to be mathematics with a narrow scope. The semantic method for settling this issue in particular cases is to perform a semantic analysis of the basic concepts of the context in dispute. If at least one such concept has a factual (nonconceptual) reference class then the context is factual – otherwise formal. This semantic analysis can be checked and supplemented by a methodological analysis: if at least one validation procedure is empirical (e.g. experimental) then the context is factual. But of course for such an empirical test to be possible the context must be factual to begin with. (Caution: computer simulations do not qualify as empirical even if they bear on factual contexts.) *Remark 3* We saw in Sec. 3.3 that singular tautologies, like ⌜Antarctica is cold or not cold⌝, can have a factual referent. It would seem then that logic is a factual context after all, contrary to our previous contention. Not so, because logic is concerned with the whole of the conceptual universe C, in particular with statements – not with their referents. (E.g. the reference class of a propositional connective is the class of statements on which it is defined.) And, as we saw in Sec. 2.1, the reference relation is not generally transitive.

Of all the factual contexts the most interesting ones, because the richest, are the factual sciences. A *factual science* is a science with a factual reference, i.e. it is a factual context capable of being tested in accordance with the scientific method. This quick characterization of a factual science is of a methodological character. A strictly semantic definition of science seems to be out of the question: a precise sense and a factual reference are necessary but insufficient to have a science, and a high degree of truth, though desirable, is neither necessary nor sufficient. For this reason we do not give here a formal definition of factual science: this is a matter for methodology (cf. Bunge, 1967a). But once we have settled for some characterization of factual science we can use our semantics to shed some light on certain aspects of the former. First of all the concept of subject matter of a scientific subject:

DEFINITION 2.25 Let $\mathscr{S}_F = \langle S, \mathbb{P}, D \rangle$ be a factual science. Then the *domain* (or *subject matter*) of $\mathscr{S}_F$ equals the factual reference class of the maximal predicate $P_{\max}$ in $\mathbb{P}$:

$$\mathit{Domain}\,(\mathscr{S}_F) = \mathscr{R}_F\,(P_{\max}).$$

Needless to say, a science need not have a homogeneous domain in the sense of Definition 3 in Sec. 2.3. And a science need have no precise deductive structure: it can be just a context, which is something slightly more structured than a mere set of statements, in that a given family of predicates is shared.

And now to the central referent of our semantics:

DEFINITION 2.26 T is a *factual theory* iff (i) T is a theory and (ii) T contains factual predicates.

We leave metamathematics, in particular the theory of theories, the elucidation of the concept of a theory. Our present concern is with the objects a factual theory is about:

DEFINITION 2.27 The *factual reference class*, or *universe of discourse*, of a theory T equals the reference class of the maximal factual predicate $P_{\max}$ in T:

$$\textit{Domain}\ (T)=\mathcal{R}_F\ (P_{\max})$$

Remark 1 Since theories in logic and pure mathematics contain no factual predicates, their factual reference classes are empty. *Remark 2* The factual reference class of a scientific theory is constituted by possibles not just actuals. For example, the genetics of a given biological species is about all possible genotypes of that species – only a tiny fraction of which are ever actualized. *Remark 3* The notion of a possible individual, or a possible factual item, is sometimes regarded as ghostly and at other times in need of modal logic. Yet it is not ghostly, for science handles it every day: thus mechanics is the study of the possible motions of possible bodies. And the concept cannot be characterized by modal logic, which is too poor a theory. But is easily defined in our semantics, namely thus: The factual item x is *possible according to the theory T* iff x belongs to the factual reference class *Domain* (T) of T. (In science we have no use for the absolute modalities handled by modal logics.)

4.2. *The Factual Reference Class of Scientific Theories*

The determination of the reference class of a factual theory is far from being a straightforward business. Scientific theories come usually unkempt: they are more or less haphazard collections of formulas accompanied by extrasystematic remarks. These hints are ambiguous and are often loaded with an obsolete philosophy, namely, operationism, which forces the observer, or experimenter, into every reference class. Our problem has a solution only if the theory is axiomatized with respect to both its form and its reference. We shall henceforth assume that the scientific theories which are the object of our analysis have been axiomatized in both respects. Not that this is necessary in order to use the concept of factual reference of a theory, but it is needed for computing the reference class with precision.

If the theory is formal, its axioms will determine the structure of the basic (primitive or undefined) concepts as well as their interrelations. For example, if one of the basic concepts of the theory is a binary associative operation on a certain set, then one of the axioms of the theory will state this much. But if the theory is factual such a formal characterization will be insufficient: it must be supplemented with semantic assumptions indicating the nature of the referents – whether they are seeds, or birds, or perhaps both. In other words, a factual theory will be a formal structure enriched with a set of assumptions concerning the referents of its basic concepts. In a nutshell, we stipulate

DEFINITION 2.28 T is an *axiomatic factual theory* iff (i) T is a factual theory (according to Definition 26), (ii) T is axiomatized and (iii) every basic (undefined) factual concept of T is explicitly assigned a definite nonempty reference class by some of the axioms of T.

This is the kind of axiomatics Hilbert and Bernays (1968) called *inhaltlich*, as opposed to *formal* axiomatics, which abstracts from the specific content of the original theory. An axiomatic factual theory is determined by both its formal (mathematical) axioms and by the nonformal (semantical) assumptions of the theory – roughly those that indicate the referents of the basic predicates in the theory. The first part of the preceding statement is a truism from the theory of deductive systems (Tarski, 1956, Ch. xii). The second part will be proved. As to the first part: calling $\mathcal{A}$ the axiom base of T, we symbolize that statement in the standard way

$$T = \mathcal{C}n(\mathcal{A})$$

which can be read: 'an axiomatic theory is the totality of consequences of its axioms'. In addition, the (nonformal) axioms of a factual theory determine not just the referents of its factual predicates but the reference class of the theory as a whole (as characterized by Definition 27). To show this we must start by construing the concept of an axiom base as the conjunction of all the axioms of the theory. We shall limit ourselves to a finitely axiomatizable theory, but infinitary logic can handle the infinite case. We stipulate then

DEFINITION 2.29 Let A_i, with $1 \leqslant i \leqslant n$, be the n axioms of a theory T. Then the *axiom base* of T is

$$\mathcal{A}(T) =_{df} \bigwedge_{i=1}^{n} A_i .$$

We can now state

THEOREM 2.1 Let $\mathscr{A}$ be the axiom base of a theory T. Then

$$\mathscr{R}(T)=\mathscr{R}(\mathscr{C}n(\mathscr{A})).$$

Proof sketch In an axiomatic theory no theorem contains predicates other than those occurring in its axiom base. In other words, every concept in an axiomatic theory is either defining (primitive) or defined. Hence deduction cannot change the kind or number of referents assigned the basic factual predicates.

THEOREM 2.2 The factual reference class (or universe of discourse) of an axiomatic theory equals the union of the factual reference classes of its axioms:

If $\mathscr{A}(T)=\bigwedge_{i=1}^{n} A_i$, then $\mathscr{R}_F(T)=\bigcup_{i=1}^{n} \mathscr{R}_F(A_i)$.

Proof By Theorem 1, Definition 9, and Definition 19.

Remark 1 The last theorem is the most important in our theory of reference. It may be called the theorem of *reference conservation* or reference invariance under deduction. It constitutes a partial justification of the dictum that the conclusions must be "contained" in the premises – a dictum that, as we shall presently see, is not generally valid. *Remark 2* The preceding theorem is far from trite: in fact it fails for nonaxiomatic theories. Indeed, in the latter new predicates, totally alien to the basic ones, can be adjoined at any step. This adjunction can be made in either of two ways: via the introduction of new assumptions or via deduction with the help of the principle of addition of formal logic. The former move is obvious: informal theories are forever under construction or repair. The second move is less obvious and deserves special notice. *Remark 3* Let t be a formula of an informal theory T. By the principle of addition, $t \vdash t \vee u$, where u is an arbitrary statement, i.e. one that need bear no semantical relation to t. In fact, as far as logic is concerned, the reference classes of t and u need not overlap, let alone coincide – i.e. t and u may fail to be coreferential. Thus t could be about stars and u about frogs – and yet u could be counted among the members of the open theory T. In general: in an open context, the reference of a consequence may be wider than that of the premises. This violates the principle of the Aristotelian organon according to which "the indemonstrable premises of a science (...) must be within one genus with its conclusions" (*Posterior Analytics*, Bk. I, Ch. 28, 88). If this principle, tacitly obeyed in most cases, is surrendered then anything can be claimed to be relevant to anything else. The result will be confusion or error or both. Such a move, hardly avoidable in informal theories (which are open contexts), is automatically precluded in axiomatic theories, which are closed contexts. While the former are semantically loose or open, hence both perfectible and corruptible, axiomatic theories are semantically tight or closed. This last expression is elucidated by

DEFINITION 2.30 A set S of statements will be said to be *semantically tight*, or *closed*, iff S has the same referents as its logical consequences, i.e. if

$$\mathcal{R}(S)=\mathcal{R}(\mathcal{C}n(S)).$$

Theorem 1 can now be restated in this way: *Axiom systems are semantically closed.* For this reason the axiomatic format, though usually unsuitable during the early construction stage, is mandatory for investigating the semantics of a theory: *it is only in axiomatic contexts that we know exactly what we are talking about.* This is because axiomatization involves fixing in advance the set of basic predicates, hence the universe of discourse, to be admitted in the given context. Thus axiomatization suffices to get conclusions that are semantically *relevant* to the premises. Should it be objected that the price to pay for semantic coherence or relevance – namely axiomatization – is too high, it may be rejoined as follows. First, price is irrelevant. Second, axiomatization is a means for attaining a number of other goals as well (see Bunge, 1973b). Third, the alternative procedure for trying to secure relevance is far more expensive than axiomatization, as it consists in adopting some system of relevant implication (cf. Anderson, 1972). This change in logic involves not only new rules of inference but also modalities – none of which seem to be called for by scientific work. And, despite commanding such a high price, this alternative procedure does not really guarantee semantic relevance, as it offers no criteria for it and it places no restrictions upon the principle of addition.

If we wish deduction to preserve reference even outside semantic contexts then we must keep a permanent watch on the consequences and screen out all those that violate semantic closure, i.e. that are not referentially germane to the premises. (One of the rare occasions on which such a precaution is taken explicitly, is in stating Craig's interpolation theorem. A standard formulation of this theorem is as follows. Let p and q be formulas such that $\vdash p \Rightarrow q$. Then there exists a third formula r, *containing only predicates occurring in p and q*, such that $\vdash p \Rightarrow r$ and $\vdash r \Rightarrow q$.) The very choice of premises in an argument should be regulated by the unspoken yet time honored *principle of relevance*: The premises should be not only mutually compatible in the formal sense but also semantically congenial. This principle of argumentation limits the range of applicability of the so called rule of augmentation of premises, according to which if p entails q then q follows also from p conjoined with an arbitrary premise r. This rule holds only if the additional premise is either idle or both formally and semantically coherent with the original premises. Once the initial assumptions have been chosen we may use a deduction technique that complies automatically with the requisite of semantical closure and has the additional advantage that it does not require guessing the conclusions beforehand, so that it may be used by a nonmathematician – e.g. a computer. (See, e.g., Hilbert and Ackermann, 1950, p. 24.) This *mechanical procedure for extracting reference preserving conclusions* is as follows. (i) Conjoin all the assumptions; (ii) expand the conjunction in conjunctive normal form, i.e. as a conjunction of binary disjunctions; (iii) detach every conjunct and every conjunction of such. The detached formulas constitute the desired set, i.e. the maximal set of consequences compatible with reference conservation. This subset of consequences is semantically closed and finite, hence a smallish subset of the infinite set of all the consequences of the given premises. We cannot have both all the infinitely many consequences of a finite set of premises, and semantical closure.

Our Theorem 2 of reference conservation allows us to determine the reference class of any factual theory cast in an axiomatic format and consequently it enables us to compare theories as to their domains. In fact the theorem backs up the following conventions.

DEFINITION 2.31 Let T and T' be two axiomatic factual theories. Then T and T' are said to be *referentially commensurable* iff their reference classes overlap, i.e. if

$$\mathscr{R}_F(T) \cap \mathscr{R}_F(T') \neq \emptyset.$$

DEFINITION 2.32 Let T and T' be two axiomatic factual theories. Then T' is said to be *referentially bulkier* than T iff

$$\mathscr{R}_F(T') \supseteq \mathscr{R}_F(T).$$

Example 1 Particle mechanics and the theory of electromagnetic fields in a vacuum are incommensurable: they are not about the same things. *Example 2* Electrodynamics and mechanics are referentially commensurable: they have common referents – bodies. *Example 3* Relativistic mechanics is referentially bulkier than nonrelativistic mechanics because, unlike the latter, it concerns both bodies and electromagnetic fields. Hence they are referentially commensurable. Consequently they are comparable – *pace* Kuhn and Feyerabend.

Remark 1 If two theories are referentially incomparable then they are not comparable at all except in strictly formal respects. In particular they cannot be compared as to explanatory and predictive powers: they do not explain and predict any shared facts. *Remark 2* The concept of referential commensurability will be supplemented with that of methodological commensurability in Sec. 5.1. *Remark 3* Our concept of referential commensurability is at variance with Kuhn's (1962). For Kuhn two theories are incomparable iff they fit different theory paradigms and thus are conceptually different even though they may concern the same things. Behaviorist and cognitivist learning theories would be incommensurable in this sense. But, since they concern the same kinds of animal, to us they are referentially commensurable and thus comparable. If they were not they could not be regarded as rival. *Remark 4* Nor does our criterion of theory commensurability agree with Feyerabend's, which is the sharing of *statements*, hence of concepts. It is of course true that relativistic mechanics "does not, and cannot, share a single statement with its predecessor" (Feyerabend, 1970, p. 82). But it is also true that both refer to bodies (cf. Example 3 above). Moreover the basic concepts of classical mechanics can be retrieved from the corresponding relativistic concepts: for example the classical mass function is a certain restriction of the relativistic mass function (see Bunge, 1970b). Therefore the two theories are commensurable, hence comparable – which is why they are being compared all the time. If this were not the case then there would be no more ground for preferring one of the theories over the other than for preferring Newton to Vivaldi. *Remark 5* The overlapping of two theories, i.e., their sharing some statements (Feyerabend's criterion) is not only unnecessary for their commensurability: it is not even sufficient. In fact two factual theories with different referents may have the same mathematical formalism, as is the case with certain theories of epidemics and of rumor propagation. It would be preposterous to declare such theories semantically commensurable even though they share infinitely many statements. What is decisive (necessary and sufficient) for theory commensurability is not the mere sharing of statements but the sharing of semantic assumptions, for these are the ones that point to the referents or their properties. *Remark 6* The moral of the preceding remarks is this. The interesting problem of comparing rival scientific theories and ascertaining their meaning diversity cannot be approached on the basis of spotty historical remarks: it requires a full fledged semantics of science and an analysis of scientific theories in the light of the former.

4.3. *Spotting the Factual Referents: Genuine and Spurious*

The preceding examination of the factual reference class of a scientific theory presupposes that the latter has been formulated in a correct way, both formally and semantically, so that finding its referents is a straightforward operation. In other words, we have assumed that the theory to be analyzed contains detailed analyses of every one of its variables: in this case Theorem 2 solves our problem. Unfortunately this situation is exceptional rather than typical: in most cases the referents of a theory are hinted at in extrasystematic remarks dropped in motivating the theory or, as we saw in Sec. 1, they are the subject of spirited controversy.

In other words, semantics can be of little help in spotting the genuine referents of a theory unless the latter has already been axiomatized in such a way that those referents are clearly exhibited. But at least semantics can help us to dig up the referents of an untidy theory and can assist us in formulating the semantic assumptions of the theory. These tasks are performed in the light of a criterion that, unlike Theorem 2, concerns unkempt theories. Here it is:

CRITERION An object is a *genuine factual referent* of a scientific theory iff it is involved in at least one of the law statements in the theory.

Otherwise, i.e. if an object falls under no law of the theory, then it is a *spurious referent* of it, no matter how many authorities may claim the opposite. For example, if a theory of electrons fails to contain law statements about instruments sensitive to electrons, then the theory does not refer to such instruments, much less to those in charge of the latter. It will not do to claim on philosophical grounds that some reference to instruments or to observers *must* be "implied" or "presupposed" by the theory, for otherwise the latter would be neither meaningful nor testable.

The preceding criterion suggests adopting the following

DEFINITION 2.33 A nonconceptual object x is a *genuine factual referent* of an informal theory T iff (i) T contains at least one law statement involving one or more variables concerning x; (ii) such law statements cease to hold upon the removal of the variable(s) in question, and (iii) T can be adjoined, without introducing any contradiction, assumptions effecting a more precise characterization of the variable(s) referring to x.

Whatever passes for referent but is not genuine is a *spurious referent*. Contemporary science contains several theories with spurious referents, i.e. that purport to but in fact do not refer to certain entities. Such theories may be called *ghostly*. We proceed to exhibit three ghosts.

Example 1 Probability is sometimes interpreted in a subjectivist or semisubjectivist way, occasionally even in the context of natural science. For instance, it is sometimes held (e.g. by Brillouin 1962) that the distribution function (or else the partition function) in statistical mechanics, and the various probability densities in quantum mechanics, concern our own information (or rather want of it) instead of either objective (though potential) properties or objective random distributions. If this contention were true then statistical mechanics and quantum mechanics would concern the cognitive subject and his mental states at least as much as autonomous physical entities. By our Criterion, if the cognitive subject were a genuine referent of those theories then these should contain law statements enabling us to explain and predict our own behavior. But they do not. Consequently the cognitive subject is a spurious referent of those theories. Equivalently: those theories are about autonomous physical systems. This conclusion is reinforced by the consideration that one learns about such systems by studying them rather than by the method of transcendental meditation. *Example 2* The so called measurement theories found in the foundations of mathematical psychology happen to be concerned with certain types of scientific concept not with the empirical operation scientists call 'measurement'. So much so that they do not involve any laws of either the measured object or the measuring instrument, whence they are of no help in designing and interpreting any measurement proper or even in estimating random errors of measurement (see Bunge, 1973c). *Example 3* It is usually held that the quantum theory of measurement (or some variant of it) accounts for actual measurements and moreover concerns the system-apparatus-observer compound. However, this theory fails to contain variables representing any properties of the observer. *A fortiori* the latter does not occur among the referents of the law statements of the theory – to the point that the author of the standard quantum theory of "measurement" acknowledged that the observer "remains outside the calculation" in his theory (von Neumann, 1932, pp. 224, 234). And although variables alleged to represent properties of an instrument do play a role in this theory, they fail to concern real instruments. As a consequence none of the formulas of this theory predicts any experimental results. A genuine measurement theory concerns a specific set-up, hence it contains specific law statements accounting for the peculiar structure and composition of the apparatus and its coupling with the measured thing. Such a specificity shows up in the constitutive equations or, equivalently, in the occurrence of non-universal parameters and constants, such as the refractive index, the electric resistivity, or the magnetic permeability. The quantum theory of "measurement" contains no such constitutive equations, hence it can refer to no genuine measurement set up. In sum, the quantum theory of "measurement" is ghostly: both the observer and the apparatus are spurious referents of it. (For details see Bunge, 1967b, Ch. 5; 1973b, Ch. 4.)

4.4. *The Strife over Realism in the Philosophy of Contemporary Physics*

The assumed referents of a scientific theory may be physical, chemical, biological, social, or what have you. Whatever their kind, the referents can be regarded in a number of alternative ways in their relation to the cognitive subject. The main doctrines that dominate the current philosophy of science in this respect can be summarized as follows.

(i) *Conventionalism* Scientific theories have no referents: they are just data summarizing and data processing tools – in particular they are prediction devices.

(ii) *Pragmatism, operationism, phenomenalism* Scientific theories are about phenomena, i.e. facts in which some cognitive subject is involved and such as they appear to him: they concern neither the subject alone (subjectivism) nor an autonomous object (realism).

(iii) *Realism* If the referents of a scientific theory exist at all then they exist on their own, i.e. independently of their being perceived or conceived: the cognitive subject creates, tests and applies hypotheses and theories instead of either creating their referents or posing as the latter.

The defense of any such stand is usually dogmatic: it consists in quoting authorities. At most it is pragmatic: it derives its strength from the heuristic power of the given philosophical doctrine, as when operationism is defended on the ground that it helps in the conceiving of experiments, or when realism is tolerated because it assists to invent "theoretical entities". Such lines of defense are weak and they foster sterile controversy. We need a different strategy, one relying on a semantical and methodological analysis of the theories concerned. We should be able to use a theory of reference to disclose the genuine referents of a scientific theory and thus evaluate the conflicting doctrines concerning such referents. How this can be done will be sketched presently.

From a semantic viewpoint conventionalism is false, since a theory with no factual reference whatever may qualify as a mathematical theory, hardly as a scientific theory. It is true that most of the formulas of a scientific theory do not involve explicitly the object variables or referents discussed in Sec. 2.4, so that the formulas look reference free, as if they were formulas in pure mathematics. But the factual reference is usually indicated or suggested by the accompanying text, sometimes called 'the prose of the theory'. And when these extrasystematic remarks are incorporated systematically into the theory, as so many semantic assumptions, the referents are brought to light. In short conventionalism, plausible as it may look in relation to the bare mathematical bones of scientific theories, is mistaken in relation to suitably axiomatized factual theories. So much for a semantic criticism of conventionalism.

The preceding semantic considerations are reinforced by the following methodological remarks. First, by refusing to make a semantic (or referential) commitment, conventionalism fails to draw the line between the various sciences and even between science and mathematics. (More on semantic formalism in Ch. 6, Sec. 3.1.) Second and consequently, conventionalism is incapable of telling us what kind of data are relevant to a given theory. Third and consequently, conventionalism does not help us to devise empirical tests, as every test of a scientific theory presupposes our knowing its reference class. Therefore the conventionalist view that scientific theories are data processing contraptions fails to do justice to both theories and data.

If conventionalism is discarded then only two serious contenders remain: phenomenalism and realism. (The truth might of course belong to a further party not involved in the current strife, but it will turn out that semantics and methodology favor just one of the current rivals.) Before the days of relativity and quanta, physics was usually interpreted in a realist way. Thus the length $L(b)$ of a bar b (or rather its length relative to a length standard) was regarded as an intrinsic and objective property of the body b. With the advent of special relativity (1905) length became a joint or mutual property or two physical systems: the body b and a reference frame f. Absolute, i.e. frame-free lengths, were swept aside in principle though not in practice. In other words, the definite description '$L(b)$' was replaced by '$L(b, f)$', short for "the length of b in (or relative to) the frame f". That is, a new length function was introduced: one with a different reference class. While the reference class of the nonrelativistic length concept was the set of physical systems, the reference class of the relativistic length concept is the union of the set of physical systems and the set of reference frames.

Now, the term 'reference frame' may be construed in either of two ways: as a physical system of a special kind, or as an observer. If the former, then nothing changes from an epistemological point of view: a property that had been regarded as absolute (or frame independent) turned out to be relative (or frame dependent). Equivalently: a relational metric property was substituted for an intrinsic one. But if reference frames are conceived as sentient beings equipped with yardsticks, clocks, mirrors, and other pieces of apparatus, then length ceases to be a strictly physical property to become a property of compound systems with a human component. And this is epistemologically revolutionary – or rather counter-revolutionary, as it reinstates anthropocentrism. Thus the expressions 'apparent length' and 'the length measured by a moving observer' were coined to indicate that the theory concerns appearances (to observers) not autonomous reality. The facts that all basic law statements are required to be frame free (or covariant), and that many magnitudes (such as the spacetime interval, the electric charge, and the entropy) are invariant, i.e. the same in (relative to) all frames, were not taken as evidence against this operationist interpretation. They were just ignored.

Two decades later quantum mechanics suffered a similar fate: its formulas were read in the light of the then dominant philosophy of physics. According to the usual or Copenhagen interpretation of quantum mechanics, every formula of the theory is about some microsystem under the action of an experimental set up controlled in an arbitrary fashion by an observer. Because of this alleged control and of the supposed docility of microsystems, which are imagined to behave as commanded by the observer, the apparatus-observer complex is usually called "the observer" – even if the whole experiment is automated. And dynamical variables are christened "observables". In this way no objective physical properties would seem to remain: they all become, just as with the operationist interpretation of relativity, observer dependent. For example, an eigenvalue is construed as a possible measurement result, an eigenstate as the corresponding observed state, a superposition of eigenstates as symbolizing our own uncertainty concerning the state of the system for not being under observation, and so forth. Consequently phenomenalism is pronounced victorious and realism dead. From here on the transition to strait subjectivism is easy: "the basic principles of physics, embodied in quantum-mechanical theory, deal with connections between observations, that is, contents of consciousness" (Wigner, 1970).

Do these non-realist interpretations of contemporary physics have any ground other than authority? Could semantics help us to make a rational choice between phenomenalism and realism? It can indeed, namely by supplying the criterion and the definition of factual reference given in Sec. 4.3. To reveal the genuine factual referents of a theory pick its basic predicates, analyze them and disclose the way they function in the central law statements of the theory – i.e. before the latter is applied or tested. Of all the alleged independent variables only those will qualify as genuine which are characterized by the theory and occur in the latter's law statements: all others make no difference and thus are ghostly, i.e. they concern spurious referents. Let us see how this criterion works in weeding out ghostly variables in a couple of important cases.

Example 1 Consider the famous Lorentz "contraction" formula

$$L(b, f)=L(b, b)\,[1-u^2(b, f)/c^2(w)]^{1/2} \tag{SR}$$

where b denotes a body (in fact any physical system), f a reference frame, and w an electromagnetic wave in a vacuum. The formula gives the length of b in (relative to) f, in terms of the length $L(b, b)$ of b relative to b itself, and of the ratio $u(b, f)/c(w)$ of two velocities: the velocity of the body with respect to the frame, and the (absolute) speed of light $c(w)$ in void. The phenomenalist construes f as an observer and interprets $L(b, f)$ as the apparent length measured by f. But, according to our criterion, this interpretation is inadmissible because the theory gives a strictly physical characterization of f, i.e. one that contains no psychological and sociological concepts. In sum, the referents of SR are b, f, and w, all three physical objects. Hence realism is upheld in relation to special relativity.

Example 2 Look at Heisenberg's indeterminacy formula:

$$\Delta_{\psi(\sigma, \sigma')}p(\sigma)\cdot\Delta_{\psi(\sigma, \sigma)}p(\sigma)\geqslant h/4\pi \tag{QM}$$

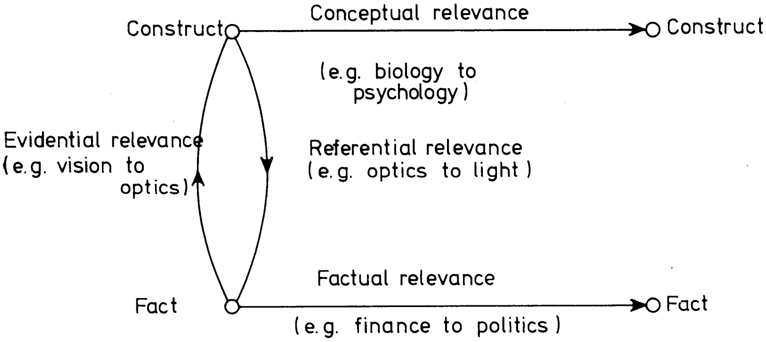
If a correct deduction from first principles is performed and analyzed, it is seen that $\psi(\sigma, \sigma')$ represents the state of a microsystem σ (e.g. an iron atom) in an environment σ' (e.g. a magnetic field) at a given instant of time. It is further realized that $\Delta_{\psi(\sigma, \sigma')}p(\sigma)$ represents the width of the momentum distribution of the microsystem σ when the system-environment complex is in the state represented by $\psi(\sigma, \sigma')$. Similarly $\Delta_{\psi(\sigma, \sigma')}q(\sigma)$ stands for the width of the position distribution of the microsystem. In short the object variables or referents of the formula QM are σ and σ'. The phenomenalist claims that σ' is the observer-*cum*-apparatus complex. However, (*a*) the formula holds even in the absence of an environment, i.e. when σ' happens to be the null individual; (*b*) the formula holds whether or not σ' includes a measurement device – for example it holds for the atoms in a star; and (*c*) the theory does not specify the properties of either observers or measurement devices: it is a strictly physical theory and a thoroughly general one, married to no particular laboratory, let alone to a particular mind. Conclusion: the phenomenalist or operationist interpretation of quantum mechanics is groundless: as with the case of relativity, that interpretation is a superstructure that the formulas of the theory fail to support.

In conclusion: there are two main ways of looking at a scientific theory. One is through a semitransparent mirror, like operationism, that exhibits the superposed images of the referent and of ourselves. The second way is by using the non-reflecting magnifying lenses of axiomatics and semantics applied to the theory itself regardless of its philosophical wrapping. The second method alone guarantees the objectivity characteristic of science. (For details see Bunge, 1973a and 1973b.)

5. RELEVANCE

5.1. *Kinds of Relevance*

The term 'relevance' is very much in vogue these days but it is seldom analyzed. In particular it is not always realized that relevance is a relation and that its relata can be of various kinds, as shown in the following diagram.


Construct
Conceptual relevance
Construct
(e.g. biology to psychology)
Evidential relevance (e.g. vision to optics)
Referential relevance (e.g. optics to light)
Factual relevance
Fact
Fact
(e.g. finance to politics)

The relevance relations involved in the preceding diagram are these:

(i) construct-construct as regards form, i.e. *formal relevance*;
(ii) construct-construct as regards reference, or *semantical relevance*;
(iii) construct-fact, or *referential relevance*;
(iv) experiential fact-construct, or *methodological relevance*;
(v) construct-action, or *pragmatic relevance*;
(vi) fact-fact, or *factual relevance.*

The last two kinds of relevance may be characterized as follows. A fact may be said to be relevant to another fact if and only if the former makes some difference to the latter. And a construct may be regarded as being pragmatically relevant to an action iff the former is part of a view or theory that is instrumental in bringing forth or preventing the given action. We shall be concerned only with the first four types of relevance. They will be characterized in the following way.

DEFINITION 2.34 Let c and c' be two constructs. Then c is said to be *syntactically relevant* to c' iff there is a context in which c is logically related to c' in such a way that c determines c' at least in part.

Example 1 In a definition, the definiens is syntactically relevant to the definiendum. *Example 2* In an argument, the premises are syntactically relevant to the conclusions. *Example 3* In a function, the independent variables are syntactically relevant to the dependent variables. If the function has an inverse, the relevance is mutual.

DEFINITION 2.35 Let c and c' be constructs. Then c is said to be *semantically relevant* to c' iff (i) c is syntactically relevant to c' and (ii) c and c' share referents, i.e. $\mathscr{R}(c)\cap\mathscr{R}(c')\neq\emptyset$.

Example 1 The specific gravity function is relevant to the weight function. *Example 2* Let c concern a gene in an organism and c' a molar or phenotypical trait of the same organism. Then c will be semantically relevant to c' just in case genetics happens to contain a law according to which c determines c' at least in part. *Example 3* The biological variables are irrelevant to the psychological ones in the context of behaviorism.

DEFINITION 2.36 A construct c is said to be *referentially relevant* to a fact (thing, state, event, process) f iff f is in the reference class of c, i.e. if $f\in\mathscr{R}(c)$.

Example Pure mathematics is syntactially relevant to science, which is in turn referentially relevant to reality. *Example 2* According to the Copenhagen interpretation of the quantum theories, the latter are relevant to the human mind, while according to the realist interpretation they are not. *Example 3* The concept of thinking (or ideation) is referentially relevant to neural activity.

We may take a further step and introduce an absolute and a comparative concept of degree of referential relevance:

DEFINITION 2.37 The degree of *referential relevance* of a theory T to a domain F of facts is equal to $\mathscr{R}(T) \cap F$.

DEFINITION 2.38 Let T and T' be two theories that are referentially relevant to a domain F of facts. Then T is *referentially more relevant* to F than T' iff $(\mathscr{R}(T) \cap F) \supset (\mathscr{R}(T') \cap F)$.

Example Biology is referentially more relevant than ethics to the facts of animal aggression, altruism, and cooperation; and political science is more relevant to war than either biology or ethics.

These concepts should not be mixed up with the methodological concept of evidential relevance, that can be elucidated by

DEFINITION 2.39 An empirical fact e is *evidentially relevant* to a construct c iff there is another construct c' such that (i) c' is syntactically relevant to c and (ii) c' is referentially relevant to e.

Example 1 The conceptual basis of the common lie detector is the hypothesis that increase in hand sweating is an index of lying. *Example 2* Dreams are (so far) irrelevant to the study of personality because there is no scientific theory in which dream contents and personality traits are related. *Example 3* Prior to the theory of evolution the differences among species were not normally regarded as pointing to (or against) the hypothesis of evolution.

In Sec. 4.2 we elucidated the concept of semantic commensurability. We can now supplement this notion with the one of methodological commensurability, which has recently come to the fore with reference to scientific revolutions. We first stipulate

DEFINITION 2.40 Let T and T' be two factual theories. Then T and T' are said to be *methodologically commensurable* iff there are empirical facts that are evidentially relevant to both T and T'.

DEFINITION 2.41 Let T and T' be two methodologically commensurable theories. Then T is said to present a *larger methodological exposure* than T' iff the set of empirical facts evidentially relevant to T includes the one evidentially relevant to T'.

Example 1 Relativistic mechanics presents a greater methodological exposure than classical mechanics because it is inextricably related to electrodynamics. *Example 2* Molecular genetics presents a greater methodological target than Mendelian genetics because it is sensitive to an additional set of evidence, namely that of the molecular structure of genes. *Example 3* Any neuropsychological theory presents a greater methodological exposure (hence more risks) than the corresponding behaviorist theory (if any) because it can claim both behavioral and neurophysiological evidence pro and con.

Now, testing a theory presupposes knowing what the theory is about – not the other way around. In other words, contrary to the verification tenet of meaning, we hold that the semantics of a theory must precede its methodology (recall Sec. 2.6). More explicitly, we lay down the methodological

THESIS 2.4 If two scientific theories are methodologically commensurable then they are semantically (referentially) commensurable as well.

The converse is false. Hence referential and methodological commensurability are inequivalent. Thus two rival theories about ghosts are singly untestable, hence methodologically commensurable in the trivial sense that no fact of experience is relevant to either. The above principle contradicts the claim – originally put forth by Kuhn and Feyerabend – that every new revolutionary theory is methodologically incommensurable with its predecessor: that is is accepted for reasons other than having been shown to be truer. On the other hand our principle fits in with the actual practices of (*a*) comparing only semantically commensurable theories, (*b*) looking for evidence relevant to the reference class shared by the rival theories, and (*c*) making rational (though perhaps wrong) choices among rival theories after having weighed their respective empirical (predictive) performances. This does not entail that factworthiness is all-important: but it is necessary.

In examining the performance and the value of alternative factual theories it is usually taken for granted that they are semantically and methodologically commensurable. That is, one uses tacitly

DEFINITION 2.40 Two factual theories are *commensurable* iff they are referentially as well as evidentially commensurable.

A final convention: Two theories may be regarded as being *semantically rival*, or *competing*, iff, being commensurable, they have different senses, i.e. if they do not 'say' the same. (Examples: rival theories of learning, or of social mobility.) If two theories had the same sense then they could differ only in the organization of the material: in this case they would be just different formulations or presentations of one and the same body of knowledge. But the concept of sense lies far ahead.

Before applying the preceding ideas on relevance we note that they are relevant to the intuitive notion of a "category mistake" (Ryle, 1949). In the light of the preceding considerations it is clear that there are no absolute category mistakes: that whether or not a given association of predicates is mistaken depends on the theory that is adopted. It is equally clear that any such mistakes are epistemic not linguistic mishaps: they do not consist in violations of linguistic rules but in departures from accepted bodies of substantive knowledge. Thus while in the context of behaviorist psychology (adopted by Ryle) the coupling of mentalistic concepts to either behavioral or neurological ones would constitute an unforgivable category *faux pas*, in either a more backward or more advanced psychological context such predicate combinations might be correct. The science of the day, not either grammar or literary criticism, is competent to judge whether any given predicate association is right or wrong. In short, category mistakes are scientific mistakes.

5.2. *The Paradox of Confirmation as a Fallacy of Relevance*

Consider the empirical generalization ⌜All ravens are black⌝, or $\ulcorner(x)\,(Rx \Rightarrow Bx)\urcorner$ for short. This statement amounts to $\ulcorner(x)\,(\neg \mathrm{Rx} \vee \mathrm{Bx})\urcorner$ as well as to $\ulcorner(x)\,(\neg Bx \Rightarrow \neg Rx)\urcorner$. At first blush, finding anything that is not a raven, such as a book, or anything black, such as a tunnel, or even that is not black, such as a carrot, would seem to count as positive evidence for the generalization. But then almost anything would so count no matter where we look. Hence experience will have little if any value: confirmation will be so cheap as to be worthless. This is one of the paradoxes of confirmation (Hempel, 1945).

This paradox, which occurs neither in real science nor in our semantics, illustrates the fallacy of irrelevance. Indeed the statement that all ravens are black happen to concern birds not anything else: this is how ornithologists understand it. Therefore to put it to the test the ornithologist will scan only the *relevant evidence*, i.e. data about birds, without bothering about nonbirds, whether black or colored. This procedure is consecrated by our semantics, in particular by Definition 2.9 (iii) and Corollary 2.4b. Indeed, the reference class of the universal conditional ⌜$(x)\ (Rx \Rightarrow Bx)$⌝ equals the reference class of the molecular predicate $R \Rightarrow B$. And this predicate is defined on the intersection of the domains of its components R and B, as stipulated in Ch. 1, Sec. 2.3. Now,

$$\text{Dom } R = \text{Birds}, \quad \text{and} \quad \text{Dom } B = \text{Things}$$

hence

$$\mathscr{R}_s\,(\text{All ravens are black}) = \text{Birds} \cap \text{Things} = \text{Birds}$$
$$\mathscr{R}_s\,(\text{Anything is either not a raven or black}) = \mathscr{R}_p\,(\neg R \vee B)$$
$$= \mathscr{R}_p\,(R \vee B) = \text{Birds}$$
$$\mathscr{R}_s\,(\text{All nonblacks are nonravens}) = \mathscr{R}_p\,(\neg B \Rightarrow \neg R)$$
$$= \mathscr{R}_p(B \Rightarrow R) = \text{Birds.}$$

Therefore only an examination of birds will be relevant to the given generalization and thus be in a position to confirm or infirm it. (The discovery of albino ravens has refuted the generalization anyway.) *Moral 1*: If semantics is to be relevant to actual science it must be consonant with the methodology of science. *Moral 2*: It is not helpful to jump from the syntax (or logic) of science to its pragmatics (or methodology) without stopping over its semantics.

Our analysis of the empirical generalization about ravens carries over to statements in theoretical science. A typical example of a simple conditional occurring in the latter is ⌜All metals are conductors⌝, or ⌜$(x)\ (Mx \Rightarrow Cx)$⌝ for short. This is a law proper for, in addition to being well corroborated, it belongs to a definite scientific theory, namely solid state theory. (For the underlying criterion of lawfulness cf. Bunge, 1967a, Vol. I, Ch. 6, Sec. 6.6.) This theory accounts for conductivity (both thermal and electrical) in terms of metal structure: it explains why metals should be conductors. In the terminology of the preceding section: M is semantically relevant to C because there is a definite context, namely solid state theory, where (*a*) M and C have common referents and (*b*) M determines C. On the other hand in ⌜All ravens are black⌝ the predicate "is a raven" is so far semantically irrelevant to "is black": at the time of writing there seems to be no confirmed scientific theory explaining the blackness of ravens in terms of any of their basic properties. In other words, in the context of present day ornithology "raven" and "black" are syntactically irrelevant to each other, and *a fortiori* semantically irrelevant as well. This situation, atypical of science, is typical of ordinary knowledge: it suggests Hume's and James' ontology of universal ontic irrelevance. Such a metaphysics can hardly avoid fallacies of irrelevance. Not so the ontology of science, which is a web of laws. But this is a different matter. In any case our semantics may be of help when the metaphysician goes astray.

6. CONCLUSION

The notion of reference, though central to the semantics of factual science, is of little interest to logicians and mathematicians, whence it has been neglected by semanticists. The only serious attempt to analyze reference in exact terms seems to have been Goodman's. Unfortunately the definition he offered was exceedingly complex and at the same time it did not require predicates to be analyzed as functions from individuals to statements. Consider: "*S* is *absolutely about* *k* if and only if some statement *T* follows from *S* differentially with respect to *k*". In turn "a statement *T* follows from *S* differentially with respect to *k* if *T* contains an expression designating *k* and follows logically from *S*, while no generalization of *T* with respect to any part of that expression also follows logically from *S*" (Goodman, 1961, p. 7). The notion of aboutness is thus made to depend upon the concepts of consequence and denotation, but the latter is left unanalyzed. If only for this reason the elucidans turns out to be more obscure than the elucidatum (Patton, 1965). No wonder Goodman's criterion has not been applied to solve any of the genuine problems of referential ambiguity listed in Sec. 1.

Our approach to the problem of reference has been direct: we have been concerned with solving the problems mentioned in Sec. 1 and with proposing a solution that could be checked. Our proposed solution has consisted, in a nutshell, in bringing to light the domain of the predicate of interest. Since the responsibility for the characterization of a factual predicate rests ultimately with a scientific theory, it is to the latter that we should turn for an identification of the entities it purports to be about. Mark: to the theories themselves rather than to any extrasystematic ("philosophical") comments about them – the more so since such comments are often philosophically biased. If somebody claims that a theory *T* is about entities of kind *K*, let him formulate *T* in such a way that *K* occurs in the domain of at least one predicate in *T*: otherwise the claim will be groundless. But even showing that a theory refers to entities of some kind does not prove that it describes or represents them. For example, a theory of the effects of food deprivation in children will refer to food without describing or representing it. But the concept of representation deserves a separate chapter: the next.

CHAPTER 3

REPRESENTATION

According to a realist semantics, scientific theories represent their referents. They constitute *conceptual representations* of actual or hypothetical bits of reality – or rather of some of their traits. The same holds for some of the predicates and some of the formulas of a scientific theory. Just for some: not all of them represent. Thus, while a distance function may represent the spacing of things, not every one of the infinitely many functions of a given distance function will represent something. Likewise not every line in a deduction within a scientific theory represents some aspect of a thing: some statements are purely mathematical. What makes some constructs representational and others not? This is, in a nutshell, the problem of the present chapter.

1. CONCEPTUAL REPRESENTATION

There is an analogy between scientific theories and pictures. In both cases the represented object may be given (rather than invented) and the representation may be more or less accurate or truthful. This analogy is the basis for the expressions 'Scientific theories depict (or portray) their referents' and 'Science mirrors (or reflects) reality'. However, these are just metaphors, hence they may not go to the heart of the matter. Indeed, the manner of representing differs in either case. First, while a pictorial representation is itself a physical object, a conceptual representation is a thing of reason. Second, while in the former case only appearances can be pictured and all else can at best be hinted at, scientific representations do not pause at the skin of reality: they aim at representing the real, which is mostly hidden to the senses and alien to ordinary experience. Third and consequently, scientific representations are symbolic (yet not metaphoric) rather than pictorial even though they may contain a few pictorial ingredients. Fourth, while a picture is bound to be interpreted (and often also felt) in as many ways as there are beholders in different moods, a scientific representation is supposed to be objective. Fifth, while the aim of the arts is to arouse or to lull, to entertain or to edify – in any event to play up on our emotions – the aim of a scientific representation is to describe its referents in a truthful way. Whatever aesthetic emotions a scientific representation may arouse derive from realizing its logical and methodological virtues or shortcomings.

In view of these important differences between pictorial representations and conceptual representations, it seems advisable to speak of *conceptual reconstructions* rather than *images* of reality. Reconstructions are constructions, artifacts resulting from hard and often ingenious work, not just impressions and images – which can be had for free. Flies have an image of reality and so have scientists, but the latter possess, in addition, what they create, namely the conceptual representations of objects that cannot be grasped with the senses. These representations are only partial and at best approximately true but they can be checked and either improved or replaced by truer ones. What is more, unlike images they are built with mathematical bricks: one more reason for not calling them 'pictures'. More on the issues of critical realism in Ch. 10, Sec. 3.3.

Not every construct represents. To begin with logical concepts are nonrepresenting even when they are referential. Thus on our theory of reference (Ch. 2, Sec. 3.2) disjunction concerns propositions. But it represents nothing. And analytic statements may refer to anything but they describe or represent nothing except logical objects: logic does not mirror reality but rather the structure of human knowledge (cf. Bunge, 1974a). Likewise with other formal objects: they too are nonrepresentational. For example the number 8 fails to represent anything. True, 8 happens to be the number of genuine planets in our solar system and thus looks as if it did represent a property of a concrete thing. Not so: 8 is not the property of any one aggregate although it may be construed as the class of all octuplets. This is why it may pop up in any context, whether formal or factual. In other words, although the number of planets equals 8,8 is not identical with the number of planets. (That is, the relation of equality involved in ⌜The number of planets equals 8⌝ is not symmetrical, hence it cannot be the one of identity. Consequently instead of writing 'Card (set of planets) = 8' we should adopt the Algol convention and write 'Card (set of planets) := 8'. Likewise to indicate that we assign x the value 8, or that we set x equal (not identical) to 8, but not *vice versa*, we should write '$x := 8$' instead of '$x = 8$'.) In any event formal constructs need not, and often cannot, represent anything – except perhaps further formal constructs, as when a point in a manifold is represented by an *n*-tuple of numbers, or when a function is represented by a series.

A neutral construct, such as a set, may represent some concrete thing or aggregate of things. Thus a portion of a continuum may be made to represent a body, and a graph may be assumed to represent an institution. But such constructs stand on their own feet and are portable from one field of inquiry to another. Hence it is not by analyzing them that we may discover what if anything they represent. This can be found out only by examining the role such constructs play in scientific theories. Such roles are sometimes assigned explicitly, namely by semantic assumptions, as will be seen in Sec. 4. In any case reference and representation are independent since nonreferential constructs, such as sets, may be made to represent while referential constructs, such as tautologies, may not represent anything. In other words it is not the case that whatever represents refers and conversely. What is true is that scientific theories both refer and represent.

Moreover not every factual statement, i.e. one with factual referents, represents facts. Consider the irreducibly negative propositions. We shall call a statement *irreducibly negative* if it cannot be transformed into an equivalent positive statement except by the trick of introducing negative or disjunctive predicates – none of which can represent properties of real things. Thus ⌜Snow is not blue⌝ is irreducibly negative although it can be converted into ⌜Snow is nonblue⌝, which is grammatically affirmative but semantically negative, since "nonblue" represents no color. Likewise ⌜There are no ghosts⌝ can be transformed into ⌜Everything is nonghostly⌝, but since "nonghostly" corresponds to no property that can be pinned down as being possessed by a thing, the statement is irreducibly negative. Our two examples illustrate a class of statements with factual reference and moreover true, but that fail to represent any facts. In general: If a proposition p represents a fact f, then $\neg p$ is just the denial of p, not a representative of non-f. (See Kraft, 1970.) In other words there are no negative facts: negation is a purely conceptual operation without an ontic correlate. We must resist the attempt to save the view that language is a picture of reality by introducing the fiction of the negative fact (Russell, 1918). The number of facts cannot be duplicated by just issuing a semantic decree. In conclusion: irreducibly negative statements do not represent anything even when true. Hence not everything that refers represents.

On the other hand every positive statement does constitute a partial representation of its referents. In particular a positive factual statement represents a fact or rather some facet of it. Thus ⌜*b* grows faster than *c*⌝ refers to *b* and *c* and represents (truly or falsely) the fact that *b* grows faster than *c*. The negate of the same proposition has the same reference class, namely $\{b, c\}$, but does not represent the "negative" fact that *b* fails to grow faster than *c*: it is just the denial of the former statement.

The distinction between reference and representation is no idle philosophical technicality: it is relevant to our understanding of topical scientific controversies. For example, biologists are still arguing about the authentic referents of the synthetic (neodarwinian) theory of evolution. So far they have produced no conclusive argument for either of the theses in dispute: that the theory is concerned with individual organisms, or with populations, or with species. However, a semantic glance at the typical formulas of the mathematical theory of evolution shows the following. First, the theory *refers* to populations or aggregates of coexisting and interacting members of a given kind (species). Hence it uses all three concepts: those of individual, species, and population. Second the theory *represents* both individual and collective traits – among other things the occasional changes in kind (speciation and extinction) occurring in a population. The failure to realize that there is no incompatibility among the three concepts, because they perform different and complementary roles, may be blamed not only on the obsolete dichotomy Platonism-nominalism, but also on the backwardness of semantics, which has never helped science find its way.

The difference between reference and representation becomes particularly clear in advanced theories such as those in physics. For example, here a probability function will *refer* to some system or some state(s) of it, while the values of that function may be taken to *represent* certain dispositions of the system – much as the mass function M refers to bodies while a particular value $M(c)$ of M will represent the mass of the body c. In quantum mechanics, every dynamical property of a system, such as its linear momentum, is represented by some operator in a Hilbert space. That is, the operator represents a property of its referent. In statistical mechanics the partition function of a multicomponent system refers to the latter but fails to represent any single property of it. It accomplishes much more: it generates the conceptual representatives of all the thermodynamic properties of the physical system. And in electromagnetic theory the value $E(\phi, x, t)$ of the vector valued function E at the field ϕ at the place x and the instant t, represents the strength of the electric component of the referent ϕ at x and t. There are infinitely many constructs in the theory, such as the powers and derivatives of E, all of them with the same referent but which represent no trait of it.

To sum up. While in factual contexts the reference relation pairs a construct off to a thing as a whole, or to a collection of things, the representation relation matches a construct with some aspect or property of the thing or collection of things. Just as we read '$\mathscr{R}cf$' as 'c refers to f', so we may abbreviate 'c *represents* f' to '$c \mathrel{\hat{=}} f$'. If c happens to be a quantitative construct we may also read '$c \mathrel{\hat{=}} f$' as 'c *represents* the strength of f', where f is a property not a whole thing or a whole fact. For example, in mathematical neurobiology the element a_{mn} of a certain matrix is assumed to represent the strength of the action (excitatory or inhibitory) of neuron m on neuron n. Table 3.1 exhibits some typical examples of conceptual representation that will guide our subsequent investigation.

2. THE REPRESENTATION RELATION

2.1. *A Characterization*

The relation $\mathrel{\hat{=}}$ of conceptual representation pairs some constructs off to some objects, whether conceptual or factual. We shall restrict our study to the case when the represented object is factual, i.e. when it is a thing or an aggregate of things, a property of either or a change in one or more properties of a system – i.e. an event. The main types of representative and their respective representees are shown in Table 3.2.

A few comments are in order. Firstly, we have included no individual constants in our list. And this because they can denote but not represent. Thus the man Socrates is denoted by his name but is represented only by certain propositions about him as well as by some definite descriptions of him. And an unspecified individual thing (an arbitrary element of a set of things) is likewise denoted by an individual variable but not represented by it. Only a collection of statements can represent an unspecified individual thing. Individual things, whether specific or unspec-

TABLE 3.1

Constructs: representing and nonrepresenting

	Construct	Represents	Refers to
Nonreferential concepts	3	—	—
	Open set of R^n	Open set of an n-manifold	—
	Region of a 3-manifold	Body or force field	—
Referential concepts	$Q(b)$	Charge of b	Body b
	$P(s)$	Probability of s	System in state s
	$6n$ dimensional manifold	Dynamical states of a system with n components	System with n components
	Partition function	—	Multicomponent system
Statements	The lake froze.	A freezing of the lake	The lake
	The lake felt cold.	Joint property of lake and subject	Lake and subject
	There are no green people.	—	People
Theories	Maxwell's field theory	Structure and spreading of electromagnetic fields	Electromagnetic fields
	Theory of evolution	Emergence, evolution and extinction of populations belonging to different species	Populations or organisms
	Theory of social mobility	Changes in occupation, social class, income bracket, or dwelling place	Groups of people

TABLE 3.2

What represents what

Representing construct	Represented object
Set of statements (e.g. theory)	System (individual or aggregate)
Predicate or structure (e.g. a set together with a relation on it)	Property of a system, relation or connection among systems
Set of singular or existential statements	Fact involving one or more systems (state, circumstance, or event)
Set of universal statements (e.g. law statements)	Pattern of a system's composition, structure, or change

ified, are best represented by logically organized sets of statements, i.e. by theories.

Secondly and as a consequence: sets devoid of structure do not represent either. Thus we may stipulate that 'E' denotes the population of African elephants, but E is incapable of representing anything. Only a structured set, such as a set together with certain functions, or a family of subsets of a given set, can represent something. And structured sets are structures, such as a semigroup or a topological space.

Thirdly, we have included existential statements among the representing constructs. It might be objected that a proposition such as ⌜There are flies⌝ hardly represents a fact. Yet it does represent a condition of the world, albeit in an imprecise fashion. ⌜There are flies in this room⌝ is a more precise statement – quite apart from its truth value. A whole sequence of such statements, each more precise than its predecessor, can be imagined. They will have different contents but every one of them, even the first statement of unqualified existence, may be regarded as representing. In sum an existential statement, if referring to some things, does constitute a partial albeit hazy representation of some facts involving those things.

Fourthly, note that we distinguish between a law statement and that which such a proposition is supposed to represent – an objective pattern. We distinguish, that is, between a law and a law statement. This distinction, emphasized by the physicist-philosopher Ampère and denied or overlooked by a number of philosophers (notably Hume, Kant, Peirce, Boutroux and their respective followers), is necessary to account for the fact that each objective pattern can be represented in a number of different ways, as will be seen in Sec. 2.2. It is also needed to explain both the process of theory construction and the history of science – which history may be regarded as a sequence of attempts to build improved conceptual representations of objective patterns or laws.

We summarize the preceding remarks in the following rather informal

DEFINITION 3.1 The relation $\hat{=}$ of *factual representation* is a relation from constructs to facts (i.e. such that $\mathscr{E}(\hat{=}) \subset C \times F$) subject to the following conditions:

(i) the properties of real things (including their interactions with other things) are represented by predicates (in particular functions);

(ii) real things are represented by sets structured by relations, functions, or operations;

(iii) facts (e.g. events) are represented by sets of singular or existential statements;

(iv) the stable (recurrent and invariant) patterns of the constitution and behavior of real things are represented by sets of universal statements.

Aside from this, as here construed $\hat{=}$ has no simple properties. In particular it is not symmetric: the represented objects do not represent back their representatives. Consequently $\hat{=}$ is not reflexive: constructs do not represent themselves. And the question whether $\hat{=}$ is transitive is pointless, since facts do not represent anything.

When the representing construct is a theory (hypothetico-deductive system) we try and be more precise, speaking of a representation function rather than just relation. In fact we propose

DEFINITION 3.2 Let T be a theory about entities of kind K and call $S(x)$ the collection of possible states (i.e. the state space) of thing $x \in K$. Further, call $S = \bigcup_{x \in K} S(x)$ the union of the state spaces of all the members of K (so that an arbitrary state of an arbitrary element of K is in S). Then T is said to be a *representation* of K's iff there exists a function $\hat{=} : S \rightarrow T$ assigning to every state $s \in S$ a statement $t \in T$. In this case '$t = \hat{=}(s)$' is read 't represents s'.

Now, a representation may be poor, fair or good, depending upon how complete and true it is. The ideal is characterized by

DEFINITION 3.3 Let $\hat{=} : S \rightarrow T$ be a representation of things of some kind. Then $\hat{=}$ is said to be *accurate* iff (i) it is bijective and (ii) $\hat{=}(s)$ is true for every $s \in S$.

As a matter of fact there are no accurate theoretical representations of real systems: every theory misses some possible states, or it includes some impossible states, or the formulas of it are only approximately true. Nevertheless one often reads that probability does constitute an isomorphic and true representation of chance events. The argument usually runs like this. Consider the possible outcomes of a random experiment such as the simultaneous flipping of two coins. Form the set E of all possible outcomes, such as getting at least one head, or not getting any head in the trial. It is easily seen that E has a Boolean structure. Now form the set T of statements describing such possible events. This set, too, will be a Boolean algebra. Hence there is a bijection $\hat{=}$ that maps E on to T in such a way that, if e_1 and e_2 are in E, then $\hat{=}(e_1)=t_1 \in T$, $\hat{=}(e_2)=t_2 \in T$, $\hat{=}(e_1 \cap e_2)=\hat{=}(e_1) \,\&\, \hat{=}(e_2)$, $\hat{=}(e_1 \cup e_2)=\hat{=}(e_1) \vee \hat{=}(e_2)$, and finally $\hat{=}(\bar{e})=\neg\hat{=}(e)$. The weakness of this reasoning is, of course, that E is not a set of actual facts but of possible facts: actual events are "positive" and "definite" (simple or composite but never alternative). Hence there is no bijection mapping actual chance events on to a theory. What is true is, that probability theory is part of the formal background of any stochastic theory furnishing a (never fully accurate) representation of some factual domain.

We postpone a fuller discussion of $\hat{=}$ to Ch. 6, Sec. 3. Here we note the relations between $\hat{=}$ and other semantic relations, as summarized in Figure 3.1.

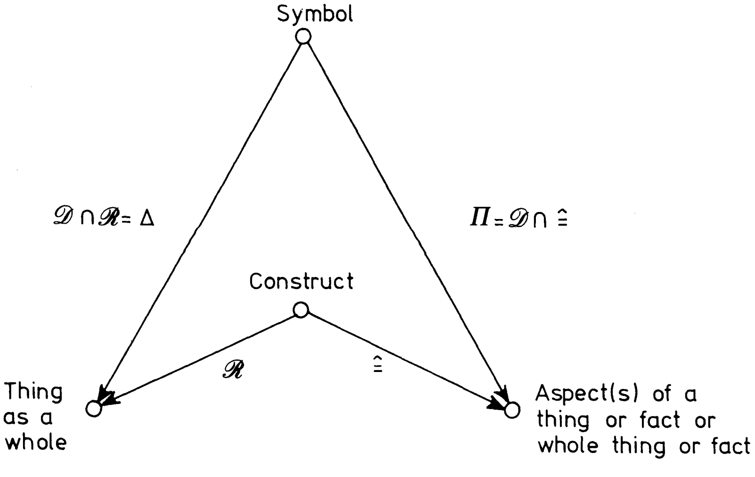

Symbol
$\mathcal{D} \cap \mathcal{R} = \Delta$
$\Pi = \mathcal{D} \cap \hat{=}$
Construct
$\mathcal{R}$
$\hat{=}$
Thing as a whole
Aspect(s) of a thing or fact or whole thing or fact

Fig. 3.1. Relations between designation (D), reference ($\mathcal{R}$), denotation (Δ), representation ($\hat{=}$), and proxying (Π).

The new relation occurring in the diagram, i.e. the intersection Π of designation $\mathcal{D}$ and representation $\hat{=}$, can be interpreted as the relation of *proxying* or deputizing. Thus we may say that, in a certain context, the symbol 'V' designates the (or rather a) velocity concept, which in turn

TABLE 3.3

Theory of the vibrating string: its main constructs and what they represent

Construct	Represents	Status
x	Projection of arbitrary point of string on Ox axis	Basic concept
l	Length of projection of string on Ox axis	Basic concept
t	Arbitrary instant of time	Basic concept
$u(x, t)$	Vertical displacement of arbitrary point of string at time t	Basic concept
T	String tension	Basic concept
ρ	String mass density	Basic concept
$(T/\rho)^{1/2}$	Velocity of travelling waves	Defined concept
$\frac{1}{2}\rho(\partial u/\partial t)^2\,\mathrm{d}x$	Kinetic energy of element dx of string	Defined concept
$\frac{1}{2}T(\partial u/\partial x)^2\,\mathrm{d}x$	Potential energy of element dx of string	Defined concept
$E =_{df} \frac{1}{2}\int_0^l \mathrm{d}x\,[\rho(\partial u/\partial t)^2 - T(\partial u/\partial x)^2]$	Total energy of string	Defined concept
$\mathscr{L} =_{df} \frac{1}{2}[\rho(\partial u/\partial t)^2 - T(\partial u/\partial x)^2]$	Excess of kinetic over potential energy densities (lagrangian density)	Defined concept
$\int_{t_1}^{t_2} \mathrm{d}t \int_0^l \mathrm{d}x\, \mathscr{L} =$ maximum or minimum	Pattern of all vibrations of string as a whole over an arbitrary time interval $[t_1, t_2]$ (global law)	Basic law statement
$T(\partial^2 u/\partial x^2) - \rho(\partial^2 u/\partial t^2) = 0$	Pattern of all vibrations of an arbitrary point of string at an arbitrary instant (local law)	Derived law statement
$u_1 = \phi_1(x - Vt)$	Wave travelling along Ox axis from left to right with speed V	Derived law statement schema (arbitrary ϕ_1)
$u_2 = \phi_2(x + Vt)$	Wave travelling along Ox axis from right to left with speed V	Derived law statement schema (arbitrary ϕ_2)
$E = \mathrm{const}$	Total energy of string remains constant in course of time	Derived law statement
$\left.\begin{aligned} u(x,0) &= f(x) \\ (\partial u/\partial t)(x,t) &= g(x) \end{aligned}\right\}$ (Initial conditions with f and g given)	Initial shape of string	Subsidiary hypothesis or datum
	Initial velocity of arbitrary point of string	Subsidiary hypothesis or datum
$u(0, t) = u(l, t) = 0$ (boundary conditions)	Ends of string fixed at all times	Subsidiary hypothesis or datum

represents speed. But we may also say that (in the same context) 'V' stands for, or proxies, speed.

We close this section with Table 3.3, that exhibits the many items of one of the simplest and most characteristic of specific scientific theories. It shows clearly that the representation assumptions are an integral part of the theory: without them the latter reduces to a mathematical formalism.

2.2. *The Multiplicity of Representations*

Representations are not unique: one and the same factual item may be represented in alternative ways. Some such alternatives are equivalent, others not. For example, a region of physical space S may be represented as a subset of a certain manifold M, which may in turn be mapped onto some subset of the collection R of triples of real numbers. Hence physical space can be represented either as a portion of M or as a portion of R^3: see Figure 3.2. These two representations are inequivalent. But in turn every region of the number space may be mapped onto some other region of the same space by means of a coordinate transformation. Since there are infinitely many possible coordinate transformations, there are infinitely many possible representations of a given region of the manifold M, hence just as many numerical representations of the original region

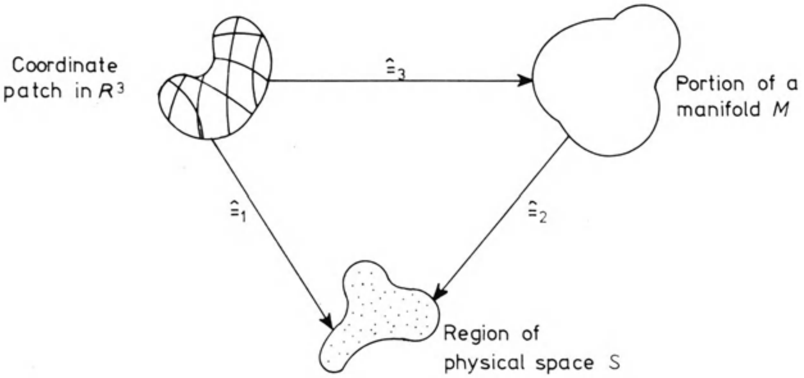

Coordinate patch in R^3
$\hat{=}_3$
Portion of a manifold M
$\hat{=}_1$
$\hat{=}_2$
Region of physical space S

Fig. 3.2. The coordinate patch constitutes a construct representing another construct, namely a region of a manifold. And each of them constitutes a *sui generis* representation of a region of physical space.

of physical space. The latter representations, i.e. the various coordinate patches mirroring one and the same region of M (or of S), are mutually equivalent. Consequently the choice among them is a matter of convenience not of truth. (Hence if meaning were dependent on truth, as conventional semantics has it, coordinate transformations should be pronounced meaningless.)

To compare alternative representations we need, at the very least, definite criteria for deciding whether any two given representing constructs represent the same factual item in different though equivalent ways. Physics abounds in such criteria, which are of a considerable epistemological and methodological interest. *Example 1* In classical mechanics any two solutions of the equations of motion, referring to the same system but to different reference frames, are equivalent representations of the same state of motion of the system provided they can be converted into one another by a Galilei transformation. *Example 2* In the relativistic theory of gravitation it is postulated that two solutions of the field equations referring to the same field and which can be transformed into one another by a continuous coordinate transformation represent the same state of the field. *Example 3* In the theory of the spinning electron every component of the electron spin is represented by an operator that is in turn representable by alternative matrices. Two such matrices represent the same spin component provided there exists a unitary transformation that carries one into the other.

Every one of the preceding criteria of equivalent representation has a definite place in some theory or other: there seems to be no theory free criterion. In any case the definitions we shall presently propose are theory bound. The first of them will concern alternative constructs in a theory, the second will handle the translation code relating mutually equivalent representations, and the third will concern equivalent theories.

The first of our forthcoming definitions hinges on the concept of basic law statement. This is a metascientific rather than a semantic concept, but we owe no apology for such an intrusion: it is unavoidable if our semantic theory is to be relevant to science. In any case the concept in question can be elucidated in the philosophy of science (Bunge, 1967a, Ch. 6). A hypothesis is called a *law statement* if and only if (i) it is universal in some respect (rather than restricted to a finite number of cases), (ii) it is systematic, i.e. a member of some hypothetico-deductive system, and (iii) it has been corroborated in some domain by scientific methods. And a proposition of this kind is called a *basic* (or *fundamental*) *law statement* of a theory T if and only if it derives from no other statement in T. We are now ready to state

DEFINITION 3.4 Let c and c' be two representing constructs belonging to a factual theory T. Then c and c' are *equivalent representations* of the same factual item [state, event or process] if and only if they can be freely substituted (i.e. substituted *salva significatione* and *salva veritate*) for one another in all the basic law statements of T, i.e. if the latter are invariant under the exchange of c for c'.

Example 1 The choice of different coordinate systems ensues in different representations of physical quantities. These representations are equivalent if they satisfy the same laws of motion and field laws. *Example 2* Let P and Q be two quantum mechanical operators representing dynamical variables. They will represent the same property of a physical system provided there exists a similarity transformation S between them, i.e. iff $Q=SPS^{-1}$. Proof: Similarity transformations leave operator equations invariant. For example, $P^2+P+I=0$, where I stands for the identity operator, goes over into $SPS^{-1}SPS^{-1}+SPS^{-1}+SIS^{-1}=0$. Calling $SPS^{-1}=Q$, we retrieve the original statement in a different notation, i.e., $Q^2+Q+I=0$. *Example 3* Let the hamiltonian of a two component system be $H=H_1+H_2+H_{12}$. Perform a canonical (unitary) transformation of the Q's and P's that "eliminates" the unperturbed energies, i.e. that makes H collapse into H_{12}. The new variables constitute the so called 'interaction representation' of the system.

Definition 4 bears on a basic aspect of concept formation in factual science, namely the representation of properties of concrete systems. Chief among such representatives are physical magnitudes such as force, stress, concentration, and temperature. Every magnitude or quantity is representable by at least one function whose values depend not only upon the physical system itself but perhaps also on the conventional units agreed upon. In other words, magnitudes and their corresponding units (a whole class of them per magnitude) must be introduced at one stroke: the set of possible units for any given magnitude should occur in the very domain of "definition" of the function. For example, in elementary electrostatics the electric force F between any two point charges is a certain function $F: B \times B \times U_F \to R^3$, where B is the set of bodies, U_F the set of force units, and R^3 the set of ordered triples of real numbers. There are infinitely many units in U_F. Every choice among them will ensue in one force value. Similarly for every other magnitude provided it is endowed with a dimension.

A way of solving this problem in a general manner is by adopting, for every scalar quantity and every component of a vector or a tensor valued quantity, the following assumption (Bunge 1971a):

AXIOM Let $A, B, \ldots, N$ be n kinds of physical system endowed with a (mutual or joint) property P. And let R designate the real number system and $\mathscr{P}(R)$ the power set of R, i.e. the family of all real number intervals. Then for every property P there exists a nonempty set U_M, called the set of *M-units*, and there is at least one function

$$M: A \times B \times \cdots \times N \times U_M \to V, \quad \text{with} \qquad V \subseteq R \quad \text{or} \quad V \subseteq \mathscr{P}(R)$$

called a *magnitude*, such that M represents P.

At first sight this assumption is redundant since, for every domain D, there are infinitely many functions "defined" on D with values in R or in $\mathscr{P}(R)$. (There are R^D such functions in the first case and $(\mathscr{P}(R))^D$ in the second.) But our axiom is not concerned with such functions except in their representative capacity. It postulates that, given any property of a system, whether known or unknown, there is at least one function that will *represent* it, i.e. that will satisfy the law statements that characterize the system. This postulate is almost in the nature of a hope: there might well exist a property not so representable. But then we would not know of its existence, for our knowledge of things and their properties *consists in* our representations of them. Furthermore the postulate does not assert that the representation of every property is unique: realistically enough the existential quantifier in it is indefinite. Different theories are likely to represent differently one and the same property. It is not up to semantics to decide which is the best representation of a given property: this is a task for science. What philosophy can do is to render explicit and systematize the criteria at work in the choice among the possible representations of a given property. One such (methodological) criterion is this: Given a property of a complex system, the best representation of it will be the one occurring in the truest and most numerous law statements concerning the given system.

2.3. *Transformation Formulas and Equivalent Theories*

To go back to the equivalent representations of a given set of factual items: how are those representations related? The answer is supplied by

DEFINITION 3.5 A statement in a theory T is called a *transformation formula* of T iff the statement relates equivalent representations of the same factual items [in accordance with Definition 4].

Example 1 The Lorentz transformation formulas relate the spatiotemporal coordinates of one and the same physical system in relation to equivalent frames of reference: they are transformation formulas of special relativity physics. *Example 2* The canonical (or contact) transformations relate different representations of the generalized coordinates and momenta of a system. Since they leave the basic (canonical) equations invariant, they are among the transformation formulas of any canonical theory.

Some remarks are in order. First, transformation formulas are not law statements and they are not data either, even though they belong in every factual theory containing spatiotemporal concepts. They just relate different representations. This point, which is obvious in the light of the preceding discussion, is very often misunderstood. Thus the Lorentz (and also the Galilei) transformation formulas are often interpreted as representing the switching on of a uniform motion – which of course would correspond to an acceleration not to a uniform relative motion. And at one time the quantal theory of canonical (unitary) transformations was regarded as the very kernel of the physical theory, whereas in fact it has no factual content of its own, being just the collection of bridges among different but equivalent representations.

We are now in a position to elucidate the notion of equivalent theoretical representation, which came to prominence in Galilei's trial and has been at the center of the realism vs. instrumentalism dispute ever since. (In fact one of the thesis of Cardinal Belarmino was that Galilei erred in holding that the heliocentric "system of the world" was true to fact while the geocentric system was false: he should have said instead that the two systems were equivalent. The same view has been advocated in recent times by conventionalists like Poincaré and by positivists like Frank and Reichenbach. However, the very notion of theory equivalence was never too clear in these debates.) We propose the following elucidation:

DEFINITION 3.6 Let T and T' be two theories with the same factual referents. Call $\mathbb{P}$ and $\mathbb{P}'$ their respective predicate bases. Then T and T' are said to be *semantically equivalent* (or to constitute *equivalent representations* of their referents) if and only if there exists a set of transformation formulas for $\mathbb{P}$ and $\mathbb{P}'$ that effects the conversion of T into T' and *vice versa*.

Example 1 Lagrangian and hamiltonian dynamics are equivalent representations of systems in general even though their formalisms are different. Indeed, there is a bridge or transformation formula between the two theories, namely $H=p\dot{q}-L$, that leaves the content invariant. *Example 2* On the other hand the geocentric and the heliocentric "systems of the world" are not semantically equivalent if only because the former has no equations of motion (but only equations for the trajectories). Only the planet trajectories, when written in geocentric or in heliocentric coordinates, are equivalent representations. Since such trajectories is all one can observe, an empiricist must conclude to the overall equivalence of the two representations. However, they are really different in every other respect: "factual" and "empirical" are not identical concepts. For example the Copernicus-Kepler-Newton representation of the solar system refers not only to the bodies in the system but also to the gravitational field that keeps them together, which the Ptolemaic representation did not. (Cf. Bunge, 1961c.)

We close this section by laying down some tenets that are germane to the preceding considerations although they belong in the pragmatics of science rather than in its semantics. Whichever their proper location here they come.

*P*1 For any factual item (thing, thing property, event) it is possible to build at least one construct that represents it.

*P*2 Given any representing construct it is possible to form at least one other construct that is semantically equivalent to the former.

*P*3 Given any representing construct it is possible to build a semantically stronger construct.

3. MODELING

3.1. *From Schema to Theory*

A single predicate, such as "round" or "competitive", can represent a trait of a complex system, never the total system. An adequate representation of a whole system, even if either is comparatively simple, requires a cluster of concepts – better, a whole theory, i.e. a body of logically interconnected statements. However, for ordinary life purposes as well as for restricted scientific purposes a mere list of salient properties is often sufficient. For example, "Brunette, medium height, 35-25-35, pretty, witty" can pass for a representation of a girl – actually a whole class of girls. Beyond this our conceptual representations of things actual or hypothetical come in all degrees of complexity and generality. It is convenient to distinguish the following kinds of representation of a system, in order of increasing complexity and generality (cf. Bunge, 1973a and 1973b).

1 *Schema* or *model object* = List of outstanding properties of an object of a given species. *Example* A neutral pion is a particle with mass 135 MeV and half life 10^{-16} sec, that decays mostly into two gamma photons.

2 *Sketch* or *diagram* = Graph of the components of an object of a given species and their functions and relationships. *Example* Flow diagram of a factory.

3 *Theoretical model* or *specific theory* = Hypothetico-deductive system of statements representing some of the salient features of a thing of a given species. *Example* A stochastic learning model.

4 *Framework* or *generic theory* = Theory representing the features common to all things of a given genus. *Example* The theory of evolution.

To put it more briefly: a schema lists items; a sketch displays in outline the relationships among the items of a schema; a theoretical model spells out the sketch; and a generic theory is a theory free from specifics but convertible into a theoretical model (specific theory) upon being adjoined a schema or model object. All four constructs are supposed to represent some real thing but each of them is necessarily incomplete as well as, at best, fairly faithful (true). These two shortcomings of our conceptual representations of the world cannot be helped except little by little and in two ways. First, by multiplying the number of conceptual representations (e.g. theoretical models) of the same object, having each of them focus on a different aspect of it: that is, by varying the viewpoint. Second, by improving on each of these partial representations. Actually this is what happens: at any given moment we have a stack of snapshots of an object, and at successive moments we have different and thicker stacks. (Proviso: that research be continued.) But enough of metaphors.

The various kinds of conceptual representation can also be characterized as follows.

FRAMEWORK (GENERIC theory) T_G
Referents: objects g of genus G
Primitive base: $\mathscr{B}(T_G) = \langle G$, Representatives of generic basic properties of the g's$\rangle$
Axiom base: $\mathscr{A}(T_G)$ = Basic assumptions "defining" $\mathscr{B}(T_G)$

THEORETICAL MODEL (SPECIFIC THEORY) T_S
Referents: objects s in species S
Primitive base: $\mathscr{B}(T_S) = \langle S$, Representatives of basic properties of the S's$\rangle$
Axiom base: $\mathscr{A}(T_S)$ = Basic assumptions "defining" $\mathscr{B}(T_S)$

SKETCH (DIAGRAM) D_S
Referents: objects s in species S
Concepts $= \mathscr{B}(T_S)$
Hypotheses: The bare bones of $\mathscr{A}(T_S)$.

SCHEMA (MODEL OBJECT) M_S
$M_S = \mathscr{B}(T_S)$.

The relations and differences between the four kinds of conceptual representation are now clearer. They can be summarized as follows:

(i) Whereas a schema is just a bunch of concepts, a sketch is a structure such as an oriented graph. A sketch includes a schema.

(ii) There is no logical difference between a theoretical model and a generic theory: both are hypothetico-deductive systems. The difference resides in their respective reference classes, and is reflected in the greater specificity of the basic assumptions of a theoretical model. Briefly, whereas $\mathscr{R}(T_G)=G$, $\mathscr{R}(T_S)=S\subset G$.

(iii) $\mathscr{B}(T_S)=\mathscr{B}(T_G)\cup M_S$, i.e. $\mathscr{B}(T_G)=\bigcap_{S\subset G}\mathscr{B}(T_S)$.

(iv) $\mathscr{A}(T_S)=\mathscr{A}(T_G)\cup H_S$, where H_S is a set of assumptions representing the specific characteristics of the members of S as outlined by the corresponding sketch or diagram.

(v) Neither is a picture of its referent: conceptual representations are all symbolic. Even scientific diagrams are symbolic and can be replaced by sets of statements. No theory can possibly resemble its referents. Thus there is no analogy whatever between a field and the differential equations representing it. Even Bohr's pictorial representation of the atom symbolizes but a small part of Bohr's model: it leaves out the equations of motion, the quantization conditions, and the resulting jump equations.

Table 3.4 displays some examples of the four kinds of construct we have just characterized.

TABLE 3.4

Some examples of conceptual representation

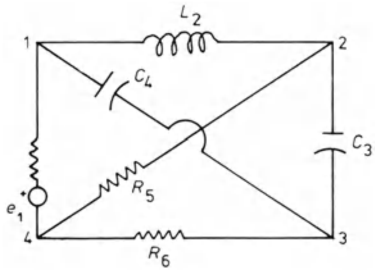
Object	Schema	Sketch	Theoretical model	Generic theory
Coin flipping	Ideal coin = ⟨Head, Tail⟩	Random sequence of Heads and Tails	Theory of Bernoullian sequences	Probability theory
Hierarchically organized group	Partially ordered set	Dominance graph or dominance matrix	Theory of dominance	———
Prey-predator system with 2 components, e.g. foxes & rabbits	Instantaneous numbers of foxes and rabbits	Numbers of foxes and rabbits in successive generations	Theory of fox-rabbit economy	Volterra-Lotka theory of predator-prey system
Corpuscle attached by spring	Mass, position, velocity, elastic force	Hamiltonian of system	Theory of the harmonic oscillator	Particle mechanics
Deuteron	Proton and neutron	Potential well	Quantum mechanics of potential well	Quantum mechanics

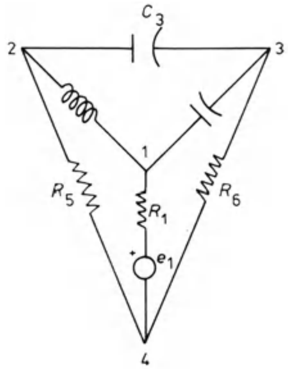
3.2. *Problems of Modeling*

The conceptual representation of things raises a number of interesting problems, some semantical, some methodological, and some technical, i.e. tractable with the resources of a special science. We can only mention a few such problems.

1 Given two representations find out whether they are *equivalent* rep-

resentations of the same system. For example, find out whether these two network diagrams are equivalent. (They are: see Seshu and Reed, 1961, pp. 1–2.)





2 Given two inequivalent representations find out whether they concern *the same* system. For example, let Δ be a block diagram of an organism. This sketch is built with objects (e.g. sets) and mappings (e.g. functions) taken from a certain category A (e.g. the category of sets or the category of topological spaces). Now consider another category B and a functor F from category A to category B. Call $F(\Delta)$ the image of the original block diagram under the functor F. This image is an alternative block diagram of the same system just in case F satisfies rather restrictive conditions: if it is faithful, regular, and multiplicative (Rosen, 1958).

3 Given a real system decide which of its traits to chop off and which ones, going beyond the data at hand, to invent. The decision will depend on the goal no less than on the available information and on the conceptual tools accessible to the theoretician: there is no unique solution to this problem because there are no recipes for building theoretical models.

4 Given a real system decide which kind of representation to build: (*a*) a black box (exogenous variables only), (*b*) a grey box (both exogenous variables and internal states), or (*c*) a translucid box (both exogenous and endogenous variables, the latter representing features of the inner workings of the system). In this case an additional factor intervenes – namely the theoretician's philosophy. Thus while positivism favors the black

box, realism encourages translucid boxes – and spiritualism cares for neither.

Finally let us list a few topical problems in the current philosophy of science involving the concept of representation.

1 Is it necessary for a successful concrete representation of some object, i.e. for a simulate, to resemble the modeled object? Undoubtedly, for otherwise the simulate would not qualify as such: it could not possibly take the place of the original in any respect. However, resemblance is required only in the respects we want to imitate. For example, a blueprint of a house must respect the relative positions of the components of the house, and so must a ball and spoke model of a molecule. Other kinds of simulate of the same objects will seize on different features, such as behavior or gross output. For example a binary digital computer bears no detailed anatomical resemblance to a brain but its net behavior is in some sort of correspondence with the functioning of a brain engaged in algorithmic thinking.

2 Does every conceptual representation have to resemble its referents? No, it need not share any properties (substantial analogy) nor need it be a homomorphic image of the represented thing (formal analogy). For example, the theory of evolution does not resemble evolution.

3 Can a theoretical model represent every trait of each of its referents? Of course not. A basic procedure of scientific theorizing, in contrast to describing, is the discarding of details and idiosyncrasies, e.g. the treatment of equivalents as if they were identicals. In other words, it is a basic methodological principle of theoretical science that equivalents (or genidenticals) be represented as being identical.

4 Are theoretical entities real or fictitious? This is an ill conceived question. First, because the "entities" concerned are not entities at all but constructs occurring in scientific theories. Second, because they may not concern entities but properties of (hypothesized) entities, such as the energy of a collapsing star. The correct question is not whether "theoretical entities" are real but whether our theoretical concepts refer to real entities and, if so, which of them represent them correctly and, if so, to what extent. And these are hardly questions that can be answered with exclusively philosophical resources: they are matters for experimental science.

5 It has been claimed, without any supporting arguments, that "what

we have called the laws of nature are the laws of our methods of representing it. The laws themselves do not show anything about the world" (Watson, 1950, p. 52). True or false? Neither: just confused. The goal of natural science is to represent nature – an objective it achieves when finding its laws. The investigation of the patterns of representation, on the other hand, belongs to psychology, epistemology and methodology. An analysis of both laws statements and the concept of representation would have avoided the confusion.

To conclude: all science aims at producing conceptual representations of its referents. (See however MacKay (1969) for the claim that the making of representations is the concern of information theory.) In the process of building such representations scientists meet challenging methodological problems that cannot be appraised if hypothesizing, modeling and theorizing are conceived as data summarizing. Whereas some such problems are peculiar to the given field of research others are undisguised epistemological problems. Philosophy can help in understanding them or at least in realizing that they involve philosophical problems.

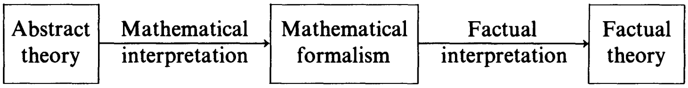
4. SEMANTIC COMPONENTS OF A SCIENTIFIC THEORY

4.1. *Denotation Rules and Semantic Assumptions*

In contemporary science theories are expressed by systems of symbols, or symbolisms. What these symbols stand for is more or less clear from (*a*) the formulas in which they occur and (*b*) the explicit designation rules that assign constructs to those symbols. An example of a *designation rule* (or proposal) is: ⌜Let '*S*' designate (name, represent, stand for, symbolize) a set⌝. Such rules are conventional in the sense that the precise symbol chosen to represent a construct is immaterial as long as every time the symbol occurs it be assigned the same construct. The resulting mathematical formalism is not an abstract mathematical theory (e.g. the general theory of groups) but an interpreted one – e.g. a theory describing a group of transformations of the Euclidean plane.

Such a mathematical *formalism* is by itself neutral with respect to matters of fact. So, unless the formalism is "read" in factual terms, it will "say" nothing about reality. Take, for example, the mathematical theory of natural selection – the hard core of the contemporary theory of evolution. As Waddington (1967, p. 14) says in colorful terms, "the whole real guts of evolution – which is, how do you come to have horses and tigers, and things – is outside the mathematical theory (...) The sheer mathematical statement is largely vacuous. The actual way it is applied, not by the mathematical theorist but by the biologist working with the subject, is not vacuous at all". This does not mean that "the whole real guts" of a scientific theory has to remain detached from its formalism: it can and must be blended with it. That is, the mathematical formalism becomes a factual theory, or rather one of a number of possible factual theories with the same underlying formalism, if the suitable factual interpretation is supplied.

The typical situation is exhibited in the following block diagram. (For a detailed analysis see Ch. 6.)



The factual interpretation of a theory is superimposed on a mathematical framework with a definite mathematical meaning and is determined by two disjoint sets of semantic rules. One is formed by the *denotation rules* or symbol-thing correspondences identifying the referents of the theory. This set constitutes what Campbell (1920, pp. 122–128) called the "dictionary" of the theory. The other is the set of *semantic assumptions* or function-property correspondences. Whereas the former point to and baptize the referents of the theory, the semantic assumptions link constructs to factual items by indicating the traits of things that the constructs are supposed, rightly or wrongly, to represent. A few examples will help getting these ideas across.

Example 1 In mechanics we find, among others, the following semantic formulas:

*DR*1 π *denotes* (names, stands for) a particle.

*SA*1 $X(\pi, f, t)$ *represents* (or measures) the position of the particle π relative to the reference frame f at the instant of time t.

Example 2 In the genetics of populations we meet

*DR*2 Let *a be* (denote, name, stand for) an allele.

*SA*2 *W dt represents* the total increase in fitness of the population of interest through the time interval *dt*.

Example 3 In mathematical sociology we come across

*DR*3 Let *k denote* a rung in a hierarchy.

*SA*3 $\sum_k k \cdot n_k$ *represents* the status of an individual with n_k subordinates on the *k*th rung of the hierarchy.

Because names are conventional, denotation rules are partly conventional. That is, the same factual items might be relabelled without qualms, as this would ensue in a different symbolism not in a new body of theory. However, inasmuch as the denotation rules indicate the hypothetical referents of the theory, they are not conventional: the theory might turn out to refer to different entities or to no real entities at all. And it is even clearer that the semantic assumptions are full blown hypotheses not just a matter of notation. They are not hypotheses about reality but about the theory-reality correspondence. Consequently a change in the semantic assumptions of a theory will ensue in a different theory with the same mathematical formalism. For example *SA*1 above is unacceptable to an operationist, who would reformulate it in terms of the measurable position values obtained by an observer attached to (or constituting) the frame *f*. As to *SA*2, it might not hold in a theory other than Fisher's mathematical theory of natural selection. Finally *SA*3 might read differently in an organization theory focusing on effective power instead of status.

The distinction we have drawn is seldom made explicitly in the scientific literature. Here one finds careless statements such as ⌜*T* is the absolute temperature⌝, which may be interpreted either as a designation rule or as a semantic assumption. It behooves the semanticist and the foundations worker to find out in each case whether the blanket word 'is' is intended to stand for "designates", "denotes", or "represents". Such distinctions are not pedantic: they make the difference between convention and conjecture. And this difference is of course of paramount methodological difference. Yet it is not made in the contemporary philosophy of science, where semantic rules and semantic hypotheses go by the common name of *correspondence rules.*

Since denotation rules are only partly conventional and semantic assumptions are wholly hypothetical, they ought not to be accepted on authority as they sometimes are. It should be possible to argue about them and even to put them to the test – though not independently of the mathematical formalism they serve. Let us try and explain. A theory constitutes no factual theory unless it includes a set of semantic formulas interpreting its basic concepts in factual terms. (Whether these semantic ingredients are displayed within the body of the theory, or hinted at in casual remarks, they are attached to the mathematical formalism.) Consequently it is the theory as a whole, i.e. the mathematical formalism together with the set of semantic formulas, that is subject to empirical tests. Fancy sending either a fleshless formalism or a boneless semantics to the laboratory.

If the empirical tests prove favorable and are themselves reliable, the formalism-semantics composite is declared confirmed until new notice. Otherwise there are three possibilities: to blame the formalism, to blame the semantic formulas, or to scrap both. The former alternative invites mending the formulas without touching the interpretation: this is a common enough procedure and one that succeeds often. The third alternative calls for a fresh start and may end up in a scientific revolution such as the founding of quantum mechanics. The middle alternative, viz., repairing the semantic formulas, seems to be less well known to philosophers but it is tried more often than radical upheavals. The following examples are well known to physicists and they show how significant a change in the semantics of a theory can be.

Example 1 The main impact of special relativity upon classical electrodynamics was forcing it to abandon all reference to the mechanical ether. From then on the theory was deemed to refer to, and represent, electromagnetic fields. Consequently all the questions concerning the properties of the ether and motions relative to it disappeared overnight. *Example 2* In the beginnings of wave mechanics the ψ function was regarded either as a real wave or as a mere mathematical auxiliary (an intervening variable). Later on its square was assumed to represent the mass density of the system associated with ψ. Eventually the so called statistical interpretation was adopted on reasonable grounds and all the rival interpretations shown to be responsible for consequences incompatible with the empirical evidence. *Example 3* Yukawa's theory of nuclear forces assumed that it was concerned with μ-mesons. But the latter refused to conform to the theory. Eventually it was found that π-mesons did satisfy the theory reasonably well. Accordingly the original semantic assumption was changed.

The semanticist cannot decide whether to change the formalism or the semantics of a scientific theory in the face of adverse empirical evidence. All he can do is to insist that the semantic formulas be formulated explicitly and clearly in order to better keep them under control. He can also alert to the philosophical concomitants of any given semantic – but this point deserves another section.

4.2. *Philosophical Commitment of the SA's*

The foregoing examples of semantic formulas (denotation rules and semantic assumptions) suggest laying down the following methodological

RULE In a well formulated scientific theory

(i) the referents should be indicated by explicit *denotation rules* bearing on the class symbols of the theory;

(ii) a denotation rule has the form: 'σ denotes a (member of the class or species) Σ';

(iii) explicit *semantic assumptions* should establish which if any property of the referent(s) does a basic predicate of a theory represent;

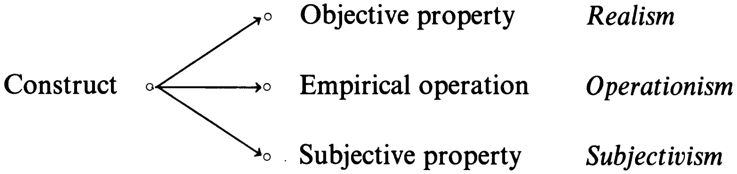
(iv) a semantic assumption has the form: 'P represents the… of a σ in Σ', where the blank names the property or relationship represented by the predicate P;

(v) the denotation of every defined symbol, as well as the representing function of every defined concept, should be consistent with the semantic formulas in which the defining items occur.

Needless to say, reference and property assignments must be consistent with the structure of the symbol or construct concerned. Thus if a single class symbol occurs in a theory then it can denote a single kind of entities rather than, say, the set of ordered pairs thing-apparatus or thing-observer. (This condition is ignored by the operationist interpretations.) And if a certain function is not time dependent then it cannot represent a change in the course of time. Secondly: it is not mandatory for every predicate in a theory to represent some property. Many predicates, particularly in sophisticated theories, represent no definite property even when (*a*) they have definite referents and (*b*) they help defining constructs that do represent. This is the case with lagrangians, partition functions, and the eigenfunctions of quantum mechanical operators other than the energy operator. Thirdly, denotation rules and semantic assumptions *sketch* factual meanings but do not exhaust them. No single piece of a factual theory, not even its semantic assumptions, gives a full characterization of the factual meaning of the theory: only the theory as a whole is fully meaningful. More on this in Ch. 5, Sec. 6.

The preceding rule can be implemented in alternative ways for any given theory. In other words, given a mathematical formalism and the above rule, there is still play enough to obtain a number of alternative factual theories – as many as alternative sets of semantic formulas be superimposed on the formalism. Which is just as well, for we want to try out different theories before we choose one of them.

Moreover, the implementation of the preceding rule depends upon one's philosophy. In particular, a semantic assumption may be construed in at least three different ways:



In the first case the theory, even should it refer to brains, would be assigned a strictly objective meaning, i.e. one independent of the cognitive subject or observer. In the second case the theory, even should it refer to galaxies far out of reach, would be assigned an operational meaning, i.e. every predicate in it would be correlated to an object-subject compositum. And in the third case the theory would concern the theoretician himself, e.g. his state of knowledge, the strength of his beliefs, and the span of his uncertainties.

These three lines in the semantic spectrum are particularly bright in relation with stochastic theories. Thus a transition probability $P(a \to b)$ from state a to state b of some system is usually interpreted in either of the following ways.

Realist SA $P(a \to b) \triangleq$ The *tendency* for the system to evolve from state a to state b.

Operationist SA $P(a \to b) =$ The *frequency* with which the system, when subject to some experimental condition, is *observed* to evolve from state a to state b.

Subjectivist SA $P(a \rightarrow b) \hat{=}$ My rational degree of *belief* in the system's transition from state a into state b.

Often either of these semantic assumptions is adopted on faith or on the strength (or weakness) of some philosophical tradition, and in any case with little concern for the structure of the concept concerned or for the role it plays in the law statements of the theory. If the probability function P occurring above is analyzed it is found that a and b denote states of some system quite apart from any experimental considerations: if there is a single object variable involved then there is no room for a second object variable such as an apparatus, much less for a third object variable representing an observer. The object variable must then represent either an external object or the theoretician (or his mind). If the former, then all the law statements in which P occurs must concern external objects whence they must be testable by handling and observing such objects. But if the referent of P is the theoretician himself, then all the statements in which P occurs must likewise refer to him (at least) and so the empirical tests of such formulas must include introspections.

We should then realize what our semantic assumptions commit us to and what their philosophical underpinnings are. And we must demand that every semantic assumption, far from being accepted on authority, be *testable* both conceptually and empirically, namely thus:

(i) A semantic assumption should *fit the structure* of the concept concerned and it should not violate any of the basic formulas in which the construct occurs. Checking this condition is of course a matter of pencil and paper: it is a conceptual test.

(ii) The construct involved in a semantic assumption should in fact describe what it is assumed to represent. Checking this condition calls for *empirical tests* of some of the formulas interpreted by the semantic assumptions.

Thus if a theory contains semantic assumptions of the operationist type, we should check (a) whether the mathematical formalism does make room for such assumptions (e.g. whether there are enough variables to represent not only the system but also the experimental set up and the observer), (b) whether the theory allows one to compute quantities representing properties of something that is not being subjected to experimental conditions, and (*c*) whether the experimental results are in fact critically dependent upon the observer and his equipment. And if the theory contains semantic assumptions of the subjectivist type, we should check (*a*) whether the theory does issue predictions about the theoretician's (or the observer's) mental states or behavior and (*b*) whether the available empirical data do bear on the cognitive subject himself rather than on objects external to him – e.g. other people. In conclusion, the semantic assumptions of a scientific theory are not conventions and they are not *dicta* beyond controversy: they are testable hypotheses. (Only, they cannot be put to the test separately from the formulas they endow a factual content with.) That they are hypotheses, and often controversial ones, can be seen from the debates on the interpretation of the mathematical formalisms of the quantum theories, to which we now turn.

4.3. *Application to Quantum Mechanics*

Quantum mechanics is probably the scientific field with the largest number and variety of semantic assumptions. (See Bunge (1956) and 'Quantum-mechanics debate', *Physics Today* **24** (1971), No. 4, pp. 36–44.) Quite often pairs of mutually incompatible semantic assumptions are nonchalantly attached to a given formula in one and the same paper. Table 3.5 exhibits a modest sample of such alternative semantic assumptions in relation to just two formulas of daily use. These formulas are

the *eigenvalue equation* $A_{op}u_k = a_k u_k$, where $A_{op} \hat{=} \text{Property } A$,

and

the *eigenfunction expansion* $\psi = \sum_k c_k u_k$.

Only a detailed analysis of the whole theory enables one to side with one or the other alternative set of semantic assumptions. An analysis performed elsewhere (Bunge, 1967b, 1967e, 1969) shows that only the realist interpretation is allowed by the mathematical formalism of quantum mechanics. We cannot go into details here but we will give a couple of reasons for rejecting the standard or operationist (or Copenhagen) version of the semantics of quantum mechanics.

A first reason why a semantic assumption cannot correlate theoretical constructs with empirical items is that the former are just not roomy enough to make place to instruments and observers. In other words, the

TABLE 5

Two sets of rival semantic assumptions (operationist and realist) for the eigenvalue equation and the eigenfunction expansion. A_{op}: operator other than hamiltonian

Symbol	Mathematical status	Operationist semantic assumption	Realist semantic assumption
A_{op}	Operator in a Hilbert space	Refers to an object-apparatus-observer block. Represents a dynamical variable of the whole.	Refers to a physical system irrespective of environment. Represents intrinsic property of system.
u_k	Function in a Hilbert space	Same referent as A_{op}. Represents state of block when measurement yields value a_k.	Refers to system (no environment). Represents nothing. (Eigenfunction expansions are mathematical auxiliaries.)
a_k	Real number	Possible value one finds upon measuring dynamical variable represented by A_{op}.	Possible value of dynamical variable A_{op}.
ψ	Function in a Hilbert space	State of observer's knowledge before performing any measurement.	State of physical system in some environment (eventually null or nonexistent).
$\|c_k\|^2$	Positive real number	Probability of getting value a_k upon measuring dynamical variable represented by A_{op}, or strength of one's belief that a_k will turn up upon measurement.	Propensity or tendency of value a_k of A

formulas do not contain enough variables that could refer to experimental situations: this is a semantic argument. A methodological reason why a semantic assumption cannot assign empirical items to a theoretical construct is that theories cannot possibly concern their own tests. Every empirical test ropes in objects other than those represented by the theory, namely pieces of apparatus. And these additional objects must be represented by additional theories. These, the auxiliary theories representing the object-experimental set-up complex, will differ according to the nature of the experimental set-up, which is rarely unique. For example, thermodynamics contains the pressure concept but no provisions for designing pressure gauges. And the latter are often employed in testing quantum mechanics but the very notion of pressure is absent from the latter. Another example: electromagnetic fields affect the growth of plants and their shape. Hence one might think of using plants as low sensitivity instruments for measuring some traits of the electromagnetic field. But it would be preposterous to claim that the electromagnetic theory, let alone quantum electrodynamics, represent plants. In general, any statements concerning the nature or the strength of the empirical evidence relevant to a scientific theory are built with the help of at least one other theory (Bunge, 1967a, 1973b). However, this is another matter – one for methodology not semantics.

5. CONCLUSION

In this chapter we have initiated an investigation of the notion of representation which will be completed in Ch. 6, Sec. 3. This notion is conspicuously absent from the conventionalist, formalist, and empiricist semantics of science. In particular the latter, which is the best known, has no use for the concept of representation because it rejects the realist thesis and holds nonlogical concepts to be ultimately percepts or logical constructions out of the latter. According to empiricism scientific theories have an observational or empirical content given them by the "correspondence rules" of the theory, which rules link the theoretical terms to observational terms, which are in turn theory-free. The notion of a real or autonomous thing, which to the realist is the object of both theory and scientific experience, does not occur at all in that doctrine.

An analysis of real specimens of scientific theory would have shown that the sole support enjoyed by empiricist semantics is tradition – a tradition of exact analyses and careful reconstructions of a nonexistent science. A case study work of the type sponsored by Törnebohm's Institutionen för Vetenskapteori in Göteborg would have shown that (*a*) the aim and outcome of theorizing is not canning sensations or even data but representing selected aspects of putatively real things, (*b*) theoretical concepts are sophisticated mathematical constructions that can neither be defined in terms of empirical operations nor construed as logical functions of observational concepts, and (*c*) empirical tests, which are conducted in order to estimate the truth value of hypotheses and theories, consist in operations planned in the light of further theories – a process that Agassi has aptly named a *bootstrap* operation.

Experience – controllable not subjective, refined not coarse – is the methodological link between theory and reality. This link does not belong to the theory: if it did then either of them would be redundant. And it is not the sole bridge that spans the chasm between theory and reality: there is also the semantic bridge constituted by the semantic assumptions of the theory. (We do not include among them the semantic conventions or designation rules, which are conventional symbol-construct links.) Those hypotheses are of two kinds:

(a) *denotation rules* of the form "Symbol s denotes thing θ";

(b) *representation assumptions* of the form "Function F represents property P of thing θ".

The semantic assumptions of a scientific theory couple symbols and their designata with supposedly real things and their properties. Since they have putatively real referents and fail to be tautologous, by Ch. 2, Sec. 4 our semantic assumptions are factual statements. Hence they have a factual content. Moreover since they are the only systematic indication concerning whatever the theory is supposed to represent, the semantic assumptions determine, at least in outline, the factual meaning of the extralogical constructs involved in them. In other words, the corrigible assumptions concerning what concepts in a theory represent which features of the world contribute to endowing a scientific theory with a factual content. Which takes us to the subject of the next chapter.

CHAPTER 4

INTENSION

⌜People make mistakes⌝ and ⌜People learn to correct some mistakes⌝ have the same referents but different senses. They represent different traits of mankind, and this quite apart from their truth values. Yet the senses of these two statements, whatever 'sense' be taken to signify, cannot be disjoint since the second proposition implies the first. The question is: What exactly are their senses and how are they related? This is the problem we face in this chapter. However, a complete solution to it will ensue only as a merger of the present chapter with the next: in this chapter only one sense of "sense" will be investigated.

1. FORM IS NOT EVERYTHING

1.1. *Concepts of Sense*

Whereas the concept of form is presumably clear, that of content is reputedly obscure. However, there is no denying that we handle contents all the time. Thus we are likely to praise logic (or to debase it, as the case may be) for being content free, hence for treating ⌜Dolphins are mammals⌝ on a par with ⌜Fractions are real numbers⌝. But it is equally clear that in order to find out the form of a construct we must know something about its content or lack of it. Change the content and the form may alter. For example, *prima facie* "longer than" is a dyadic relation but, when analyzed in relativistic physics, it proved to be triadic: here we must write ⌜x is longer than y relative to z⌝.

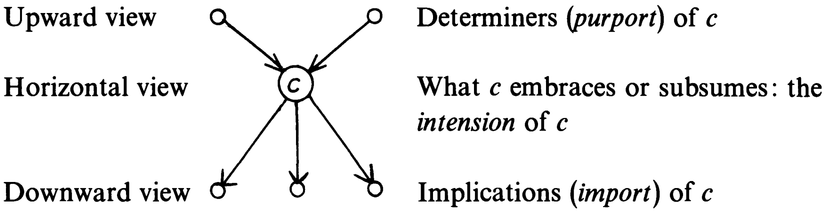
Scientific constructs come with both a form and a content. These two aspects can be distinguished but not separated: only logic can afford to deal with pure form. But distinguished they must be if we wish to understand what makes logic different from the rest, and if we are to explain the recurrence of the same mathematical structures in different fields. Take for instance "fitter than" and "more intelligent than". The two are isomorphic in the sense that they are different interpretations of the relation $>$, and so they have different contents. Even coextensive concepts, i.e. concepts with the same instances, may have different contents. For example, "mammal" and "hairy" are isomorphic and coextensive, yet they have ostensibly different contents. Likewise "avian" and "feathery", "the successor of 1" and "the smallest prime number", and so on and so forth. In short, a scientific construct is not characterized by its form and extension alone.

Granted then that, outside logic, there is something, variously called 'content', 'sense', 'intension', or 'meaning', that is not to be written off. Granted also that reference and extension are of little if any help in determining contents, as coreferentials and even coextensives, such as "human" and "cruel", can have different senses. Granted, in sum, that contents or senses are *sui generis* objects, distinct from forms, reference classes, and extensions, and thus far more difficult to pin down than either of these. The question is: What do we do with those elusive objects? Two attitudes are possible: retreat or attack. The former is argued for in this way: "Senses have always been obscure. They have defied the best philosophical minds. Therefore they are hopelessly obscure. Consequently we had better give up any attempt to clarify them". The outcome of this policy is to allow the beast to roam in the wilderness of obscurantist philosophy: after all, "man is a sense-making animal" (Quine, 1966, p. 175), and nothing human is reputedly alien to philosophers. We refuse to adopt the defeatist stand. We propose to launch an attack with a view to taming the wild beast. Our motto will be "Divide and conquer".

In fact our point of departure is recognizing that there is not a single concept of sense but rather three different concepts. We shall show later on that they are related, but we begin by distinguishing them as so many dimensions of sense. Indeed we shall propose and explain the thesis that a construct may have the following kinds of sense:

(i) the totality of its conceptual determiners, or *purport*:
(ii) the set of constructs it subsumes or embraces – its *intension*;
(iii) the totality of its *implicates*, or *import*.

In each case the sense of a construct is a set of constructs. (Note the difference with respect to both reference and extension: a scientific construct will refer to, and extend over, sets of factual items.) In each case the content of a construct is logically related to the construct itself. (Note again the difference with respect to both reference and extension.) But in the case of purport we look upwards, at the impliers; in the case of intension we look horizontally, into the construct; and in the case of import we look downwards, at the consequents:


Upward view
Determiners (*purport*) of *c*
Horizontal view
c
What *c* embraces or subsumes: the *intension* of *c*
Downward view
Implications (*import*) of *c*

Example The concept Δ of Euclidean triangle.

Purport (Δ) = The (finite) set T of statements in Euclidean geometry sufficient to characterize all the properties of Δ.
Intension (Δ) = $\{\Delta$ is a closed plane figure with 3 rectilinear finite sides$\}$.
Import (Δ) = $\{\Delta$ is a polygon, The inner angles of Δ add up to π, etc., etc.$\}$
Referents (Δ) = The (infinite) set of plane figures.
Extension (Δ) = The (infinite) subset of plane figures satisfying the conditions T of triangularity.

Of these three concepts of sense that of purport seems to be new although it sounds quite natural. It will be shown to embrace the concept of gist, or basic purport, of everyday use. The second concept of sense, viz., "intension", seems to elucidate the *significatio* of Medieval logicians as well as the *compréhension* of the Port Royal *Logique* (Arnauld and Nicole, 1662). The latter was translated by Sir William Hamilton as *intension*, a term that has been with us – and often against us – ever since (see Kneale and Kneale, 1962). A similar concept occurs in Bolzano (1837), Meinong and many others as *Inhalt*. Frege used it under the label of *Sinn* (sense) though without subjecting it to an exact theory. And, at least formally, our concept of intension is close to the *L-content* of Carnap (1942). Finally the third dimension of sense, namely import, is near the *C-content* of Carnap (1942), the *intension* of Lewis (1944, 1951) and of Castonguay (1971), and the *truth content* of Popper (1963b, 1966).

Our three concepts of sense have then roots in common sense or in technical philosophy. However, our theories of them exhibit important differences with respect to previous theories. For example, our theory of import does not involve any notion of truth: it is a purely syntactic theory employing the algebraic theory of filters. Besides, our calculus of import joins with our views on the nature of a factual theory, to yield new ideas on factual content as well as on meaning changes. Moreover we shall show that the three concepts – those of purport, intension, and import – are interrelated. (For example, purport and import are mutually dual in the algebraic sense, and in the case of axiomatic systems the intension is included in the union of purport and import.) Finally it will be seen that none of our concepts of sense can be equated with that of semantic information, and that they must also be kept separate from the concepts of testability.

Since we shall deal systematically with three different concepts, we may expect one theory for each. In point of fact we shall evolve the three theories independently and shall explore their relations at the end. We shall find it convenient to start at the middle, i.e. from intension. Purport and import will be tackled in the next chapter. But before proceeding to build a theory of intensions we must show why we need one and we must guard against confusing 'intensional' with 'intentional', 'nonextensional', and 'non truth functional'.

1.2. *Extension Insufficient*

The set theoretic and model theoretic treatments of mathematical concepts have been so effective, and they have had such a clarifying and unifying power, that they have buttressed the belief that only extensions (domains of individuals) matter. This belief may be spelled out in the following tenets:

E1 Every mathematical entity is either an individual or a set.

E2 Reference is the same as extension.

E3 Senses or intensions are either ghostly or reducible to extensions.

Criticism of E1 This thesis is particularly apparent in the theory of relations as treated in set theory. Here it takes the form: "Every relation may be defined simply as a set of ordered *n*-tuples". This claim is exaggerated: the predication relation in logic (i.e. the relation between a predicate and its subject or subjects, as in $\ulcorner Pa \urcorner$), the membership relation in set theory, and the relation of satisfaction in model theory, are not definable as sets of ordered couples. Hence although sets, in particular the extensions of relations, are extensional objects, their theories are not purely extensional. What is true is that, in *most* cases in *pure* mathematics (not so in factual science), only the extension of a relation *need* be taken into consideration for purposes of *proof*. In any case it is incorrect to identify a relation with its graph or extension. Bourbaki, for one (or rather many), draws a careful distinction between a relation and its graph (extension), in particular between a correspondence (e.g. a function) and its graph. A correspondence between the sets A and B is defined as a certain triple $F=\langle G, A, B\rangle$, where G is the graph of F (Bourbaki, 1970, Sec. 3). Finally, it is not true that all mathematical theories are reducible to set theory – nor, *a fortiori*, that they can dispense with intensions. The basic concepts of category theory, namely those of object and morphism, are not so reducible. It is rather the other way around: the set theoretic predicates may be defined in categorial terms (Lawvere, 1966). In sum, the extensionalist program of reducing every mathematical object to either an individual or an aggregate of individuals subject to the postulates of (some) set theory, though remarkably successful, is not fully feasible.

Criticism of E2 To begin with, while both predicates and propositions can be assigned referents, only predicates are normally said to have an extension. And even in the case of predicates a distinction between extension and reference class is needed. For example,

$$\mathcal{R}\,(\text{Prime number}) = \text{Natural numbers},$$
$$\mathcal{E}\,(\text{Prime number}) = \text{Prime numbers}.$$

The difference is even more pronounced outside mathematics. Thus

$$\mathcal{R}\,(\text{Flying on a broomstick}) = \text{Witches},$$
$$\mathcal{E}\,(\text{Flying on a broomstick}) = \emptyset.$$

A more respectable example is this:

$$\mathcal{R}\,(\text{Rational}) = \text{Mankind}, \ \mathcal{E}\,(\text{Rational}) \subset \text{Mankind}.$$

Finally an example from recent science: while the (hypothetical) reference class of "parton" is the set of partons, nobody knows what the actual extension of that theoretical predicate might be, not even whether it is nonvoid. In conclusion, *Reference* $\neq$ *Extension*. And, as we shall see in Ch. 9, Sec. 1.1, in general extensions are not subsets of reference classes: only the extensions of unary predicates are.

Criticism of E3 The thesis that senses can be written off altogether is false even for set theory, the paradigm of a "purely extensional theory". Indeed set theory distinguishes a predicate from its extension. So much so that it devotes an entire axiom (schema) to the relation between a predicate and its extension, namely the principle of abstraction or its improved version the principle of separation (*Aussonderung*). According to this axiom schema, to every well formed formula Px of the lower predicate calculus and to every set A there corresponds another set B whose elements are exactly those members x of A for which Px holds (cf. Suppes, 1960, p. 21):

$$(\exists B)\,(x)\,(x \in B \Leftrightarrow x \in A \;\&\; Px).$$

In our terminology A is included in the *reference class* of P while $B = \mathscr{E}(P)$ is the *extension* of P. In the case of monadic predicates, like the one involved in the previous formula, $\mathscr{E}(P) \subseteq \mathscr{R}(P)$. *Example.* Let $A = \mathbb{Z} =$ The set of integers, and $Px = \ulcorner x$ is prime$\urcorner$, with $x \in \mathbb{Z}$. Then $\mathscr{R}(P) = \mathbb{Z}$ and $\mathscr{E}(P) =$ The set of prime numbers $\subset \mathbb{Z}$. The axiom of separation is not a device for getting rid of whatever sense a predicate may have: it just clarifies the *relation* between a predicate (whichever its sense may be) and its extension. Finally, the extensionalist program fails altogether in factual science. For example, the predicates "metallic", "good thermal conductor", and "good electric conductor" have the same extension although their senses are obviously different. If only for this reason the predicates occurring in theoretical science cannot be defined *within* set theory, i.e. in the last analysis solely in terms of membership in some set, as Suppes (1967, 1969) has been advocating untiringly. We also need semantic assumptions concerning reference and representation, as we have seen in Chapters 2 and 3.

To sum up, the extensionalist thesis is false even within mathematics: we may refrain from looking at intensions for a while but we cannot suppress them. Nor should we mistake them for other characters, such as psychological intentions and modalities. This confusion deserves a separate subsection.

1.3. *'Intensional': Neither Pragmatic nor Modal*

Principia Mathematica (Whitehead and Russell, 1927, pp. 72ff and 659ff) misled several generations of philosophers by its quaint use of the word 'intensional'. What *PM* meant by this term was not one of the *semantic* concepts of sense but a certain class of *pragmatic* concepts, often called 'propositional attitudes', "such as what somebody believes or affirms, or the emotions aroused by some fact". Typical "intensional" terms would be 'doubts', 'is puzzled by', 'believes', 'knows', and 'asserts', which designate relations between a person and a proposition. Correspondingly a statement of the form $\ulcorner x$ believes that $y\urcorner$ is often called an *intensional context.* And any attempt to clarify and systematize such pragmatic statements is often said to belong to *intensional logic.*

To find out what such misnomers designate let us examine a typical example of an "intensional context". Consider the statements

$$p = \ulcorner\text{Milk is good}\urcorner \tag{1}$$

and

$$q = \ulcorner\text{Baby believes that } p\urcorner. \tag{2}$$

Whereas p is allegedly a perfectly "extensional" construct, q would be an "intensional" one, for its truth value is not a function of the truth value of the subordinate statement p alone. If it were, the truth value of q would remain invariant upon replacing p by any of its equivalents – which is not the case. Indeed, p is equivalent to

$$r = p \,\&\, (s \vee \neg s) \tag{3}$$

which, substituted for p in (2), yields

$$t = \ulcorner\text{Baby believes that milk is good and } (s \vee \neg s).\urcorner \tag{4}$$

which, unlike q, is false.

All this has nothing to do with intensions although it might be claimed to concern intentions. In fact, since tautologies are void of content, p and r above have the same intension. But they fail to have the same effect on all persons: whereas to a logician the utterance of r conveys just as much information as p does, to a layman r may sound more informative than p and to a baby r is likely to sound as gibberish. (More on information in Sec. 3.2.) All of this is pragmatics not semantics.

The actual situation is simply this. The statement q in (2) depends not just on the statement p but also upon a person, namely Baby, and a circumstance or time t. In brief, $q = B(\text{Baby}, p, t)$ where 'B' stands for the belief function. In general, pragmatic statements involving propositions are of the form: $q=P(\text{person}, p, t)$, where 'P' stands for a pragmatic function such as that of knowing. No wonder logic cannot cope with such hybrids. Little wonder that the truth value of such a statement q depends not just on the truth values of its subordinate statement p but also on the background knowledge and the state of the person concerned. In other words P is not a truth function – more accurately, it is not a truth preserving map. P is a *non truth preserving pragmatic function* – or, if preferred, an inten*t*ional function. To call it 'inten*s*ional' is to equivocate.

Pragmatic functions constitute a subclass of the class of non truth preserving functions. Some of the latter may not involve any person in an overt way. For example, ⌜*p* is conceivable⌝, ⌜*p* is doubtful⌝, ⌜*p* is puzzling⌝, ⌜*p* is beautiful⌝, ⌜*p* is suggestive⌝, ⌜*p* is useful⌝, and ⌜*p* is provable⌝ are not ostensibly pragmatic statements. Yet they can be construed as covert pragmatic statements. For example, ⌜*p* is conceivable⌝ may be construed as short for ⌜There is at least one person who can conceive of *p*⌝. But this is irrelevant to semantics. What is relevant to it is what all such statements have in common. And this is that they resist the substitution of equivalents. More explicitly: In all of these cases the subordinate proposition may not be replaced by any of its logical equivalents *salva veritate*, i.e. with preservation of the truth value of the total proposition. To paraphrase *PM*: A function of a construct will be called a *truth preserving function* when its truth value with any argument is the same as with any formally equivalent argument. Otherwise it will be called a *non truth preserving function* – by no means an "intensional function".

Modal statements are of course the best studied of all those that violate the principle of the substitutivity of equivalents. Since they, too, fall outside ordinary ("extensional") logic, Lewis and Carnap concluded (probably misled by the unorthodox use of 'intensional' in *PM*) that they must obey some systems of "intensional logic". Thus we have a first confusion: "Whatever is modal is intensional". Now indulge in the most common fallacy and obtain "Whatever is intensional is modal". These two mistakes engender the now rather popular misconception that modal logic is the key to whatever boxes ordinary logic cannot open: not only possibility but also intensions and "propositional attitudes". This conflation of "modal" with "intensional" can only vitiate semantics.

We shall keep clear of these confusions and will make no use of that all purpose tool, modal logic. The very idea of an "intensional concept", as opposed to an extensional one, is a misfit: if a concept has an intension it also has an extension even if nil. (The converse is false: sets have no intensions.) Hence the expressions 'intensional function' and 'intensional context' are misnomers. Frege's name for them, namely *ungerade* (oblique), and Quine's 'referentially opaque', are preferable. However, since 'referential opacity' has at least three different significations (Sharvy, 1972), we prefer the longer but nonmetaphorical and more explicit name 'non truth preserving'. And we shall gladly leave the study of non truth preserving functions (or inten*t*ional concepts) to pragmatics, epistemic logic, modal logic, and any other discipline that be deemed competent to handle them.

We now turn to a study of intensions conceived as semantic objects.

2. A CALCULUS OF INTENSIONS

2.1. *Desiderata*

Our aim is to build a calculus of intensions enabling us (*a*) to exactify the notion of intension and (*b*) to compute the intension of a complex construct, such as a conjunction, out of the intensions of its components. And we wish our calculus to formalize and articulate the following intuitive desiderata:

D1 All and only predicates and statements shall have an intension.

D2 The intension of a construct shall be a set.

D3 The intension of a construct in a context $C = \langle S, \mathbb{P}, D \rangle$ shall be comprised between $\emptyset$ and C, i.e. if $P \in C$, then $\emptyset \subseteq \mathcal{I}(P) \subseteq C$.

D4 If P and Q are both either predicates or statements, then

(*a*) $\mathcal{I}(P \,\&\, Q) \supseteq \mathcal{I}(P), \mathcal{I}(Q)$

(*b*) $\mathcal{I}(P \vee Q) \subseteq \mathcal{I}(P), \mathcal{I}(Q)$

D5 If $\mathcal{I}(P) \subseteq \mathcal{I}(Q)$, then

(*a*) $\mathcal{I}(P \,\&\, Q) = \mathcal{I}(Q)$

(*b*) $\mathcal{I}(P \vee Q) = \mathcal{I}(P)$

D6 If two constructs are identical so are their intensions.

D7 Tautologies are intensionally minimal and contradictions are intensionally maximal.

D8 The intension of $\ulcorner P \& (P \Rightarrow Q)\urcorner$ is the same as that of $\ulcorner P \& Q\urcorner$.

D9 Cointensives are coextensive – but not conversely.

D10 The smaller the intension of a construct the larger its extension.

We proceed to exhibit an axiom system that satisfies all these desiderata. (For interesting alternatives see Leonard (1967), Suszko (1967), Castonguay (1972), and Weingartner (1973).)

2.2. *Principles and Definitions*

Intension shall be construed as a certain function that assigns to every element of a conceptual universe of discourse a certain subset of the latter. More precisely, we lay down the following axiomatic

DEFINITION 4.1 Let U be a set of predicates or of statements. The *intension* function $\mathscr{I}$ is the function from U onto the family $\mathscr{P}(U)$ of subsets of U such that, for any P and Q in U,

(i) If $P \& Q$ is defined, then $\mathscr{I}(P \& Q) = \mathscr{I}(P) \cup \mathscr{I}(Q)$;
(ii) $\mathscr{I}(\neg P) = \overline{\mathscr{I}(P)}$;
(iii) If $P = Q$ then $\mathscr{I}(P) = \mathscr{I}(Q)$.

Example 1 $\mathscr{I}$(Female human) $= \mathscr{I}$(Female) $\cup\, \mathscr{I}$(Human). *Example 2* $\mathscr{I}$(Unattached) $= \overline{\mathscr{I}(\text{Attached})}$. *Example 3* Since $\ulcorner 1+2=3\urcorner$ is the same proposition (though not the same sentence) as $\ulcorner 3=1+2\urcorner$, $\mathscr{I}(1+2=3) = \mathscr{I}(3=1+2)$. Note that, pragmatically, the two sentences are different: while the first expresses the unique outcome of a correct addition operation, the second summarizes one of the possible results of a correct decomposition of 3.

Remark 1 The generalization of the first axiom to multiple conjunctions, whether finite or infinite, is immediate:

$$\text{If } \textstyle\bigwedge_i P_i \text{ is defined, then } \mathscr{I}(\bigwedge_i P_i) = \bigcup_i \mathscr{I}(P_i) \text{ for } P_i \in U,\ i \in N.$$

Remark 2 Definition 1 tells us, so to say, how intensions behave not what they are. Indeed, it fails to prescribe how to find the intensions of the constituents of the compound constructs. (Likewise truth functional logic shows us to compute the truth value of a compound statement given the truth values of its components.) The problem of determining the basic intensions will be faced in the next Chapter. *Remark 3* The converse of clause (iii) in Definition 1 is false. Thus ⌜Temperature $(b)=t$⌝ and ⌜Temperature $(b')=t'$⌝, though cointensive, are not identical (unless of course the referents b and b' happen to be the same).

We can now elucidate certain interesting derivative concepts.

DEFINITION 4.2 A construct is said to be *intensionally empty* iff its intension is equal to the empty set.

DEFINITION 4.3 A construct in U is said to be *intensionally universal in U* iff its intension equals U.

DEFINITION 4.4 If P and Q are in U, then P is said to be *intensionally included* in Q (or intensionally poorer than Q) iff $\mathscr{I}(P) \subseteq \mathscr{I}(Q)$.

DEFINITION 4.5 Two constructs in U are said to be *cointensive in U* iff their intensions are the same: $\mathscr{I}(P)=\mathscr{I}(Q)$.

DEFINITION 4.6 Two constructs are said to be *intensionally independent* ("perpendicular") iff their intensions are disjoint:

$$P \perp Q =_{df} \mathscr{I}(P) \cap \mathscr{I}(Q) = \emptyset .$$

DEFINITION 4.7 Two constructs are called *intensionally dependent* iff they are not independent – i.e. if their intensions overlap.

Before looking at the consequences of our assumptions and definitions let us note that an intensionally empty construct is a construct all right – not a nonconstruct. For example tautologies will prove to be intensionally void (Theorem 5). The intension function $\mathscr{I}$ is defined for predicates and statements and for them only, hence sets and nonconstructs fail to have a definite intension. One such nonconstruct results from the conjunction of predicates defined on disjoint domains, such as "wind" and "stupid". Our axiom (i) does not apply to "stupid wind" (Yeats) because there is no nonvoid domain on which "stupid wind", or any other such freak, is defined. In other words, for a predicate to have a definite intension, even a null one, it must be well constructed to begin with:

recall the rule in Ch. 1, Sec. 1.3. This does not entail throwing poetry away but just noting that some poetic expressions, though pragmatically significant, are semantically nonsignificant.

2.3. *Main Theorems*

Firstly an immediate consequence of Definition 1:

COROLLARY 4.1 For any constructs P, Q, R such that $P = Q \,\&\, R$ is defined,

$$\mathscr{I}(P) \supseteq \mathscr{I}(Q), \mathscr{I}(R).$$

Example By definition, $x < y = x \leqslant y \,\&\, x \neq y$. Since "$\neq$" equals "not $=$", $\mathscr{I}(\neq) = \overline{\mathscr{I}(=)}$ is nonempty. Hence $\mathscr{I}(<) = \mathscr{I}(\leqslant) \cup \mathscr{I}(\neq) \supseteq \mathscr{I}(\leqslant)$. Briefly, "$\leqslant$" is intensionally poorer than "$<$".

THEOREM 4.1 If P and Q are either predicates or statements, and $P \vee Q$ is defined,

$$\mathscr{I}(P \vee Q) = \mathscr{I}(P) \cap \mathscr{I}(Q).$$

Proof By logic $P \vee Q = \neg(\neg P \,\&\, \neg Q)$. By Definition 1(iii), $\mathscr{I}(P \vee Q) = \mathscr{I}(\neg(\neg P \,\&\, \neg Q))$. By (ii), the RHS is $\overline{\mathscr{I}(\neg P \,\&\, \neg Q)}$. By (i) the last formula equals $\overline{\mathscr{I}(\neg P) \cup \mathscr{I}(\neg Q)}$. Finally, by (ii) we get $\overline{\overline{\mathscr{I}(P)} \cup \overline{\mathscr{I}(Q)}}$ which, by one of de Morgan's theorems in the algebra of sets, gives the desired result.

Example Elucidate possibility of the conceptual kind as $\Diamond p =_{df} p \vee q$ with q indeterminate. Then $\mathscr{I}(\Diamond p) = \mathscr{I}(p) \cap \mathscr{I}(q) \subseteq \mathscr{I}(p)$. That is, possibility statements are intensionally included in their "extensional" bases – which is one reason for avoiding them (whenever possible!).

COROLLARY 4.2 If $\mathscr{I}(P) \subseteq \mathscr{I}(Q)$, then $\mathscr{I}(P) = \mathscr{I}(P \vee Q)$.

Proof By the algebra of sets $\mathscr{I}(P) \subseteq \mathscr{I}(Q)$ iff $\mathscr{I}(P) \cap \mathscr{I}(Q) = \mathscr{I}(P)$. Apply Theorem 1 to the LHS of the last equation.

Remark This result shows that the converse of axiom (iii) is false: cointensiveness does not imply identity. For, if it did, we should have $P = P \vee Q$. This suggests that our concept of intension does not coincide with that of sense. We shall see in Ch. 5 that the former is included in the latter.

THEOREM 4.2 If P and Q are either predicates or statements, then

$$\mathscr{I}(P \Rightarrow Q) = \overline{\mathscr{I}(P)} \cap \mathscr{I}(Q) = \mathscr{I}(Q) - \mathscr{I}(P) \subseteq \mathscr{I}(Q).$$

COROLLARY 4.3 If P and Q are either predicates or statements, then

$$\mathscr{I}(P \Leftrightarrow Q) = \mathscr{I}(Q) - \mathscr{I}(P) \cup \mathscr{I}(P) - \mathscr{I}(Q) \equiv \mathscr{I}(P) \mathbin{\Delta} \mathscr{I}(Q)$$

where 'Δ' designates the symmetric difference or Boolean sum.

We shall use this result in Sec. 2.6 in relation to differences in intension. It will turn out that "$\mathscr{I}(P) \mathbin{\Delta} \mathscr{I}(Q)$" is a good measure of the intensional difference between P and Q. Hence while the extension function $\mathscr{E}$ obliterates the differences between equivalents, the intension function $\mathscr{I}$ brings those differences into the open: see Figure 4.1

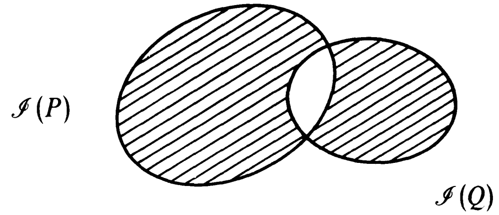


Fig. 4.1. The intension of an equivalence equals the (symmetric) difference between the intensions of the members of the equivalence (Corollary 3).

THEOREM 4.3 For any two predicates (or statements) P and Q, if $P \Rightarrow Q$ is defined then

$$\mathscr{I}(P \,\&\, (P \Rightarrow Q)) = \mathscr{I}(P \,\&\, Q).$$

Proof By Definition 1(i) and Theorem 2,

$$\begin{aligned}\mathscr{I}(P \,\&\, (P \Rightarrow Q)) &= \mathscr{I}(P) \cup (\overline{\mathscr{I}(P)} \cap \mathscr{I}(Q)) = \\ &= (\mathscr{I}(P) \cup \overline{\mathscr{I}(P)}) \cap (\mathscr{I}(P) \cup \mathscr{I}(Q)) \text{ by distributivity.}\end{aligned}$$

The first factor equals U and its intersection with the second factor equals the latter, which by Definition 1(i) in turn equals $\mathscr{I}(P \,\&\, Q)$.

THEOREM 4.4 For any predicate (or statement) P in U

(i) $\mathscr{I}(\neg\neg P) = \mathscr{I}(P)$;

(ii) $\mathscr{I}(P \vee \neg P) = \emptyset$;

(iii) $\mathscr{I}(P \,\&\, \neg P) = U$.

A generalization of Theorem 4(ii) is

THEOREM 4.5 Tautologies are intensionally void.

Proof Consider an arbitrary construct P and a tautologous one T. Because the collection of intensions is an algebra of classes, intersections determine greatest lower bounds or infima. Hence in particular $\inf\{\mathcal{I}(T), \mathcal{I}(P)\} = \mathcal{I}(T) \cap \mathcal{I}(P)$. By Theorem 1 the RHS equals $\mathcal{I}(T \vee P)$. Since the least intension is $\emptyset$, we have $\mathcal{I}(T \vee P) = \emptyset$. And, since the intension of P is comprised between $\emptyset$ and U, we conclude that $\mathcal{I}(T) = \emptyset$.

An obvious consequence of this theorem is that logic cannot alter contents. More precisely, we have

COROLLARY 4.4 Tautologies, if conjoined, do not add; if disjoined they deplete. I.e., for any predicate (or statement) P in U and any tautologous construct T of the same rank,

(i) $\mathcal{I}(P \,\&\, T) = \mathcal{I}(P)$
(ii) $\mathcal{I}(P \vee T) = \emptyset$

Proof By Definition 1 and Theorems 1 and 5.

COROLLARY 4.5 Contradictions, if conjoined, blow up; if disjoined they change nothing:

(i) $\mathcal{I}(P \,\&\, \neg T) = U$
(ii) $\mathcal{I}(P \vee \neg T) = \mathcal{I}(P)$.

Proof Similar to the previous proof.

COROLLARY 4.6 Repetition is futile:

$$\mathcal{I}(P \,\&\, P) = \mathcal{I}(P \vee P) = \mathcal{I}(P).$$

Consequently 'P', '$P \,\&\, P$' and '$P \vee P$' are not different concepts but different signs representing the same concept.

Finally a couple of results concerning the relations in intension between constructs that hold logical relations.

THEOREM 4.6 Whatever entails contains: If P entails Q, then P contains Q in intension and conversely:

$$\vdash P \Rightarrow Q \quad \text{iff} \quad \mathscr{I}(P) \supseteq \mathscr{I}(Q).$$

Proof First the left to right implication. By Theorem 2, $\mathscr{I}(P \Rightarrow Q) = \overline{\mathscr{I}(P)} \cap \mathscr{I}(Q)$. But by hypothesis the conditional is analytic and, by Theorem 5, analytic conditionals are intensionally empty. That is, $\overline{\mathscr{I}(P)} \cap \mathscr{I}(Q) = \emptyset$, which amounts to $\mathscr{I}(P) \supseteq \mathscr{I}(Q)$. Now the converse, i.e.: If $\mathscr{I}(P) \supseteq \mathscr{I}(Q)$ then $\vdash P \Rightarrow Q$. Assume the antecedent. Then, since $\mathscr{I}$ is by hypothesis a one to one map, there is a construct R such that $P = Q \,\&\, R$. And by logic $P \vdash Q$.

COROLLARY 4.7 If two constructs are analytically equivalent then they are coextensive and conversely:

$$\vdash P \Leftrightarrow Q \quad \text{iff} \quad \mathscr{I}(P) = \mathscr{I}(Q).$$

Contingent biconditionals, whether formal or factual, do not enjoy the same privilege – hence nonanalytic equivalents are not everywhere substitutable *salva significatione*. For example, in Euclidean geometry the propositional schemata "x is an equilateral triangle" and "x is an equiangular triangle" are equivalent but not analytically so – hence the two predicates are not cointensive. There would be little point in keeping the two concepts, as we do, if they were cointensive. As noted with reference to Corollary 3, the intension of an equivalence consists in whatever is not common to the equivalents. Coextensiveness is a bridge not a gate. This is a reason for not couching definitions, which are identities, in the form ⌜A iff B⌝. We shall return to this question in Ch. 10, Sec. 2.2.

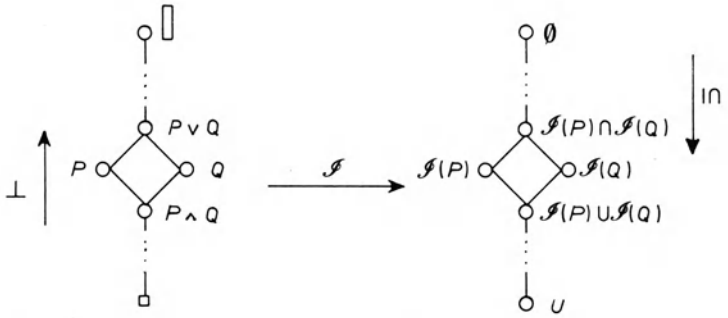


Fig. 4.2. The lattice of predicates and the lattice of intensions: anti-isomorphic.

Definition 1 and the preceding theorems and their corollaries show that there is a *duality*, or anti-isomorphism, between constructs and their intensions, in the sense that $\mathscr{I}$ carries meets into joins and conversely. Look at Figure 4.2.

2.4. *Intensional Difference and Family Resemblance*

The poorer the intension of a construct the more general it is. In other words, the more general (or less specific) of two constructs is the one with the smaller intension. Thus "atom" is more general than "helium atom" precisely because it is the weaker of the two. The difference in intension between a species concept S and a genus concept G, i.e. their *differentia specifica*, consists in whatever characterizes S but not G. Since in our semantics intensions are sets, it is natural to exactify the notion in this way:

$$\delta(S, G) = \mathscr{I}(S) - \mathscr{I}(G) \equiv \mathscr{I}(S) \cap \overline{\mathscr{I}(G)}.$$

The generalization of this concept to an arbitrary pair of constructs calls for the notion of symmetric difference, or Boolean sum, of sets. In this way we assure that the richer or stronger construct will show through. More explicitly, we introduce

DEFINITION 4.8 Let P and Q be either predicates or propositions. Then the *difference in intension* between them is the symmetric difference between their individual intensions:

$$\delta(P, Q) = \mathscr{I}(P)\ \Delta \mathscr{I}(Q) \equiv \mathscr{I}(P) - \mathscr{I}(Q) \cup \mathscr{I}(Q) - \mathscr{I}(P).$$

Example Let T and T' be two axiomatized theories with axiom bases (taken conjunctively) A and $A' = A$ & B respectively. T' is the stronger of the two theories, i.e. the one with the richer intension, whereas T is the more general (less specific). For example, T could be the general theory of groups and T' the theory of Abelian groups: in this case the extra axiom B would be the commutative law. The difference in intension between the two theories is $\delta(T, T') = \mathscr{I}(A \& B)\ \Delta \mathscr{I}(A) = \mathscr{I}(B) \cap \overline{\mathscr{I}(A)}$. This is the newness in intension, relative to A, contributed by the additional axiom. This enrichment in sense is balanced by a shrinking in coverage or extension: $\mathscr{E}(A \& B) = \mathscr{E}(A) \cap \mathscr{E}(B) \subseteq \mathscr{E}(A)$.

Let us apply the newly won concept to the case of independent or "perpendicular" constructs characterized by Definition 6:

THEOREM 4.7 The difference in intension between two intensionally independent construct equals the union of their separate intensions:

$$\text{If } P \perp Q, \text{ i.e. } \mathscr{I}(P) \cap \mathscr{I}(Q) = \emptyset, \text{ then } \delta(P, Q) = \mathscr{I}(P) \cup \mathscr{I}(Q).$$

Proof By Definition 8 and recalling that, if A and B are disjoint sets, then $A \cap \bar{B} = A$.

Example Aristotle's metaphysical categories (substance, quality, time, etc.) are intensionally disjoint. Hence any two of them differ in whatever both connote.

Corollary 3 can now be reworded as

THEOREM 4.8 The intensional difference between two constructs equals the intension of their equivalence:

$$\delta(P, Q) = \mathscr{I}(P \Leftrightarrow Q).$$

Proof By Corollary 3 and Definition 8.

Although equivalences obliterate extensional differences, they bring out differences in intension. This was to be expected in view of the anti-isomorphism between constructs and their intensions noted at the end of Sec. 2.3. Indeed, if the logical difference between two constructs is exhibited by their inequivalence, their semantical difference must be given by their equivalence. In other words, the function δ defined by Definition 8 does the opposite of its counterpart, the function $d: U \times U \to U$ that takes pairs of constructs into their inequivalences, i.e. such that

$$d(P, Q) = (P \,\&\, \neg Q) \vee (P(\neg P \,\&\, Q) = P \not\Leftrightarrow Q \text{ for } P \text{ and } Q \text{ in } U.$$

Furthermore any mapping $f: U^n \to U$ preserving all the values of the *logical* distance d may be regarded as a *tautological transformation*. On the other hand any mapping $g: U^n \to U$ that preserves all the values of the *semantical distance* δ may be interpreted as a transformation (whether logically valid or not) that does nothing to bring the constructs either closer together or farther apart. We shall not make a detailed investigation of the algebraic structure given to U by δ in analogy to the one induced by d on U. (For the latter see Blumenthal and Menger 1970.) For our purposes it will be enough to prove

THEOREM 4.9 Let U be a Boolean algebra of either predicates or statements. Then the family of their intensions, i.e.

$$\mathbb{I}=\{\mathscr{I}(P) \mid P\in U\}=\{\mathscr{I}(P)\in\mathscr{P}(U)\}$$

is a ring of sets to be called the *ring of intensions* of U.

Proof To begin with $\mathbb{I}$ is not empty: even if all the P's in U are tautologies, $\mathbb{I}$ contains the empty set. Next, by Definition 1, the union of any two intensions is an intension – namely the intension of a conjunction. Likewise for the differences. Consequently the intersection of two intensions is an intension. Hence $\mathbb{I}$ is a ring of sets.

This theorem is just a formal prerequisite for the following investigation, which is of interest to semantics. The set valued function δ is a non-numerical measure of the "distance" in intension between two constructs. This is no metaphor: we shall presently show that δ is formally similar to the distance function in a one dimensional quasimetric space. (While in a metric space two points are the same just in case they are not separated, in a *quasimetric* space distinct points need not be separated. In these spaces the weaker condition $\delta(x, x)=0$ holds instead of $\delta(x, y)=0$ iff $x=y$.) We shall call *pseudoquasimetric space* any nonempty set equipped with such a metric δ. We proceed to prove that intensions form such a space.

THEOREM 4.10 Let $\mathbb{I}$ be a ring of intensions. Then the structure $\langle\mathbb{I}, \delta\rangle$, where $\delta:\mathbb{I}^2\to\mathbb{I}$ is the difference in intension, is a pseudoquasimetric space, i.e. the function δ satisfies all of the following requisites:

(i) $\delta(P, Q)\supseteq\emptyset$;
(ii) $\delta(P, Q)=\delta(Q, P)$;
(iii) $\delta(P, Q)\,\Delta\delta(Q, R)=\delta(P, R)$;
(iv) $\delta(P, P)=\emptyset$

for any P, Q, R in the substrate U of $\mathbb{I}$.

Proof All four properties but the third are easily checked by recalling Definition 8. The "triangle equality" (iii) follows from the associativity of the Boolean sum and the equations: $A\Delta A=\emptyset$ and $A\Delta\emptyset=A$.

The interest of this theorem is twofold. First, it reinforces our decision to regard "$\mathscr{I}(P)\ \Delta\mathscr{I}(Q)$" as the difference or distance in intension between P and Q. We may now picture intensions as points on a line: see

Figure 4.3. Second, we can now define neighborhoods,

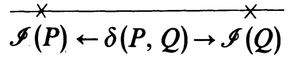

$\mathscr{I}(P) \leftarrow \delta(P, Q) \rightarrow \mathscr{I}(Q)$

Fig. 4.3. The intension space is one-dimensional

that will prove to be both mathematically and philosophically significant:

DEFINITION 4.9 Let $\mathbb{I}$ be a ring of intensions over an algebra U of constructs and call A a fixed subset of $\mathbb{I}$. Then for P in U we call

$$N_A(P) = \{\mathscr{I}(Q) \mid Q \in U \quad \text{and} \quad \delta(P, Q) \subset A\}$$

the *A-neighborhood* of P.

This concept of neighborhood in the space of intensions elucidates the fuzzy notion of *family resemblance* which the second Wittgenstein and his followers made so much of. Indeed, the A-neighborhood of a construct P is the collection of intensions such that their distance to the intension of P falls within the given intension interval A. Every $N_A(P)$ is thus a collection of constructs near P. The smaller A the closer the kinship.

Furthermore the neighborhoods introduced by Definition 9 generate two different topologies. More precisely, we can prove

THEOREM 4.11 The collections of neighborhoods

$$\mathscr{B}_1 = \{N_A(P) \mid P \in U \quad \text{and} \quad A \subset \mathbb{I}\}$$
$$\mathscr{B}_2 = \{N_A(P) \mid P \in U \quad \text{and} \quad A \in U\}$$

each constitute a base for a topology on U. That is, in each case (*a*) the union of neighborhoods equals U; (*b*) if a construct P belongs to the intersection of two given neighborhoods, then there is a neighborhood of P that is included in the intersection.

By virtue of the first clause of this theorem, no construct in a given universe of discourse is isolated, and the various families of constructs cover the entire universe. By virtue of the second clause any construct belonging to two families can be placed in a third family included in the overlap of the former and therefore constituting a set of close relatives. We shall not pursue this line here although it sounds promising because topology – rather than natural language – is the natural tool for refining concepts of vicinity and togetherness.

In closing we note that the semantic concept of family resemblance is not reducible to pragmatic (e.g. psychological or linguistic) concepts. Thus the semantic distance between two constructs is determined by the theory in which they occur not by the way they are conceived or mentioned by a subject under certain circumstances. In particular, semantic distance is unrelated to both psychological association and linguistic correlation. Whether or not two constructs are semantically close in a given context, they may be strongly correlated for some subjects and uncorrelated for others, particularly if they have never thought of them before. This being so, the attempts to account for semantic sense in psychological or in linguistic terms are bound to fail.

3. SOME RELATIVES – KINDRED AND IN LAW

3.1. *Logical Strength*

Our concept of intension is coextensive with that of logical strength. In fact, by Theorem 6 the stronger a construct the richer it is and conversely: $P \vdash Q$ iff $\mathcal{I}(P) \supseteq \mathcal{I}(Q)$. Moreover our calculus of intensions satisfies a number of conditions characterizing the notion of logical strength or deductive power. In particular, the desiderata D4, D5, D7, D8 and D9 listed in Sec. 2.1 are fulfilled by the two notions. In turn, the concept of logical strength coincides with that of specificity: the more specific (less generic) a construct the stronger it is. Thus the concept of reading skill is logically stronger and intensionally richer than the generic concept of skill. Consequently a theory of reading skill would be more specific (stronger) than a general theory of skills.

Although the concepts of strength and of intension are coextensive, they are not cointensive. (Logical strength is cointensive with import, to be studied in Ch. 5, Sec. 3.) In fact our Definition 1, which determines the intension of "intension", does not contain the concept of deducibility, which is on the other hand the very kernel of the concept of logical strength, and which defines both Carnap's and Popper's concepts of strength. This situation exemplifies our thesis that coextensives need not be cointensive or, for that matter, not even coreferential. A simpler example is this: ⌜Ormazd is unhappy iff Ahriman is happy⌝. A more sophisticated example: "consistency" and "satisfiability" (or "validity in a model") are coextensive, whence finding a model of a set of formulas proves the latter's consistency; nevertheless the syntactic notion is not cointensive with the semantic one.

We now turn to a few other notions that are often mixed up with that of intension.

3.2. *Information*

Intension is often equated with information content. Thus it is frequently asserted that, whereas synthetic propositions convey information, analytic formulas do not. No doubt there is a relation between intension and information: the greater the former the richer the latter. However, the two are not identical. To begin with a construct *has* a content whereas a signal, such as a written sentence, *conveys* (carries, transmits) information *to someone*. In other words, whereas intension is defined on the set of constructs, information is defined on the set of pairs signal-subject, where a subject is a viewer or hearer competent to decode the signal. Information, to be received, requires a receiver, i.e. a system equipped with suitable receiving and decoding devices. The sentence '$(-1)^2 = 1$', which expresses or conveys the statement that the square of minus one equals unity, transmits no information to a baby and it may convey the wrong information to a mathematician under the influence of LSD. In short, whereas the concept of intension is semantical, that of information is pragmatical: the former is subject free, the latter is tied to a subject.

Upon shoving aside the all important information receiver and the no less important information channel, it is possible to build an impersonal or semantic concept of information. In fact, upon removing the concept of a subject from the foregoing considerations and rendering them more explicit, we are left with the following principles.

INF 1 If a signal (mark, sign, inscription, sound, etc.) is a sentence or represents a sentence, then the *information conveyed by the signal* is the proposition designated by the sentence.

INF 2 If S and S' are sets of signals representing the sets of propositions P and P' respectively, then

(i) the information conveyed by S is *larger than or equal to* the one conveyed by S' iff $P \supseteq P'$;

(ii) the *information gain* accompanying the substitution of S for S' equals $P - P' \equiv P \cap \overline{P'}$.

These three propositions define what will be called our *semantic concept of information*. It boils down to this: The information or message conveyed by a signal consists of the proposition or propositions the signal stands for. It follows (*a*) that nonpropositional signals convey no information, (*b*) that the greater the content of a proposition the richer the information carried by the signal representing that proposition, and (*c*) that the truer a proposition the more accurate the information carried by the signal representing that proposition. These platitudes may be baptized *Simple Simon's semantic view on information*.

Simple Simon's view contrasts with the quantitative theories of semantic information proposed over the past two decades, from that of Bar-Hillel and Carnap (1953) to Hintikka's (1968, 1970). In particular, Simple Simon's view is at variance with the former, which initiated what is today a bulky body of literature on the subject. This theory boils down to identifying the amount of information conveyed by a statement s with its improbability:

$$cont\ s = 1 - Pr(s).$$

(For a similar proposal see Popper, 1963b.) Once this formula is accepted the whole of the elementary theory of probability can be rewritten in terms of the *cont* function and exploited for semantic purposes. The outcome is as tidy as the probability calculus, of which it is a rewrite – but it is, alas, hollow. Indeed the *cont* function can hardly be said to exactify any intuitive concept of content. For one thing contradictions turn out to have maximal content, while alternations such that the probabilities of the disjuncts are complementary are assigned zero content. For another the theory assigns coextensives the same probability, hence the same content, and is therefore at a loss to distinguish ⌜The grass is green⌝ from ⌜The weather has been mild and wet⌝, which do not even have the same referent. Worse, the theory is not applicable, because neither it nor inductive logic (with which it is sometimes associated) tell us how to assign probabilities to the constituent statements – nor even how to interpret the expression '$Pr(s) = p$'. Whence there is no point in claiming that every formula, for example the Schrödinger equation, conveys a definite amount of information lying between 0 and 1. Thou shalt not take the name of the Number thy God in vain.

Other quantitative theories of information are paraphrases of, or heavily indebted to, the much talked about but little understood mathematical theory of communication, or statistical information theory (Shannon and Weaver, 1949). The problem this theory set out to solve was to identify and measure the information carried by physical signals distorted by random perturbations causing noise. The basic notion in this theory is that of the probability of a random binary physical signal – not of a proposition or statement. This probability is a number that can be estimated provided something is known or assumed about the source-channel-receiver-environment system. It is, let us emphasize, the probability of occurrence of a *random physical event*, not the probability of a construct, let alone an orderly one such as a scientific hypothesis. The theory postulates that the amount of information carried by such a signal s_i, which pops up with probability p_i at the receiving end of a communication channel, depends only on this probability, namely thus:

$$I(s_i) = -\log_2 p_i \text{ bits.}$$

Consequently the average amount of information H conveyed by a string of N mutually independent random signals is given by Shannon's formula

$$H = \sum_{i=1}^{N} p_i I(s_i) = -\sum_{i=1}^{N} p_i \log_2 p_i \text{ bits/signal, with } \sum_{i=i}^{N} p_i = 1.$$

If all N signals are equally probable then $p_i = 1/N$, whence $H = \log_2 N$ bits/signal. In information theory H is often interpreted as a measure of the *unexpectedness* or *news value* of the message. Or, what is equivalent, the amount of information measures the ignorance or uncertainty of the receiver concerning the signal about to be received. All these are pragmatic or psychological concepts not semantic ones. Not even Kolmogoroff's generalization of Shannon's formula is relevant to semantics, as his information theoretic "entropy" is defined only for random variables, i.e. for variables associated with some probability distribution.

The preceding information theoretic concepts are far removed from the semantic concepts of intension and content. For one thing the concept of randomness, central to any application of probability theory, makes no sense with reference to propositions. (The idea of stumbling on a given proposition with a certain probability does make sense but only in the context of pragmatics – and in any case bears no relation to the content of the proposition.) For another the information theoretic concept of amount of information is in a sense the very opposite of the semantic concept of content, inasmuch as a signal conveying a proposition with a definite and rich content, such as a law of conservation of energy, has zero information content for a physicist, as it comes as no surprise to him. For all these reasons the statistical theory of information is irrelevant to semantics.

In view of the fundamental difficulties met by the quantitative theories of semantic information utilizing the statistical information theory, we shall stick by Simple Simon's concept of information. However we find no use for it precisely because it is a semantic concept, whereas information and communication belong in pragmatics. Whether or not a sign carries information to someone depends on the receiver's background knowledge and thirst for new knowledge as much as on the content of the statement being transmitted. Thus a tautology new to a given subject is informative to him, whereas a factual sentence he knows by heart is not. In any case neither the semantic concept of information nor any of the pragmatic concepts of information can be equated with the concept of intension.

3.3. *Testability*

Peirce, Frege, the Vienna Circle and the first Wittgenstein held the identity of meaning and verifiability. Carnap transformed this thesis into the equation of meaning with test procedures: "the meaning of a sentence is in a certain sense identical with the way we determine its truth and falsehood; and a sentence has meaning only if such a determination is possible" (Carnap, 1936). This thesis was subjected to such devastating criticisms (see, e.g., Williams, 1937; Russell, 1948; and Hempel, 1965) that it has now been almost abandoned by philosophers. (Quine, 1971, is about the only faithful left.) However, it survives among scientists and even, in a watered-down form, in Popper's thesis that "the degree of testability of a statement increases with its content" (Popper, 1963a, 1963b). *Prima facie* this thesis is plausible enough: tautologies have no content and are insensitive to empirical tests, whereas factual statements have a content and are presumably susceptible to such tests. Moreover, a disjunction is both intensionally poorer and less vulnerable to experience than either of its disjuncts, and "It is possible that *p*" says less and therefore risks less than "*p*".

It is difficult to assess the above thesis in the absence of a full fledged theory of degrees of testability (not just *a posteriori* confirmation) in relation to content – not to speak of a theory of content. However, the thesis is ruined by the following counterexamples. A singular observational statement such as "That apple is red" is both intensionally indigent and eminently testable. On the other hand any of the partial differential equations of physics has such a rich content that it is not fully testable. In fact one can check only a few of its (infinitely many) solutions at some of the argument values. Of course such a rich statement has more opportunities of being tested and clashing with data than a scientifically trivial statement such as "That apple is red", which can be checked at a glance. That is, the richer the intension of a statement the more tests it *calls for* in order to find out its truth value, i.e. the greater the target it offers to experience. But this can hardly be called 'testability'. It might be more aptly be called *un*testability: the more a statement "says" (the richer its sense) the more there is to be checked, hence the less fully can this task be performed. This capacity of a statement to come to grips, or making contact, with experience (Popper's 'testability') we shall call *empirical sensitivity* (or target or cross section). The empirical sensitivity of a statement (or of a collection of statements) is, intuitively speaking, inversely related to its testability proper, or capacity of being checked for truth.

This is not the place to elaborate on this problem. Therefore let me express telegraphically my views on the relations between sense and the methodological concepts we are discussing.

(i) The concepts of sense, empirical sensitivity, and testability, are *heterogeneous*, hence not mutually reducible. The first is semantic whereas the other two are methodological. The determination of sense is a theoretical affair and moreover it precedes the determination of both empirical sensitivity and testability. Besides, the degree of testability of a statement must be assessed in relation to a host of items, both theoretical and empirical, in the background knowledge.

(ii) The *empirical sensitivity* (target, cross section) a statement offers to data increases with its sense.

(iii) The *testability* of a statement is inversely related to its sense and increases with its systemicity, i.e. with the strength of the ties it holds to other members of a body of knowledge. (A strictly stray statement would be untestable. A statement belonging to a theory that partially overlaps with another theory may have the chance of being double checked: directly and via the other theory.)

(iv) The *degree of confirmation* (empirical validation, corroboration) of a testable statement decreases with its sense. Unlike empirical sensitivity and testability, which are estimated prior to tests, degrees of confirmation are assigned, if at all, *a posteriori*. And they bear no fixed relation to either empirical sensitivity or testability; in particular, a highly testable statement may prove utterly false. On the other hand degrees of confirmation depend critically upon test procedures: different techniques will corroborate (or infirm) a hypothesis to different degrees.

4. CONCLUDING REMARKS

Our theory of intensions boils down to the calculus in Sec. 2. This calculus enables one to clarify a number of obscure notions, such as those of intensional inclusion and intensional independence. We shall see in Ch. 9, Sec. 1.6, that our theory also allows us to state and even prove the reciprocal relation between intensions and extensions. Moreover, it enables us to compute the sense of a whole as a function of the senses of its parts – provided we either know the latter or do not care too much for them, for remaining satisfied with finding intensional relations. This shortcoming will be remedied in part in the next chapter, where we shall learn to find the full sense of a theoretical construct. It will turn out that the intension of a construct is included in its full sense.

Our theory construes intensions as basic or irreducible semantic objects. It is couched in set theoretic terms but it does not reduce intensions to either extensions or reference classes. Nor does our theory make use of modal concepts. In particular, our Definition 4 is at variance with the following construals of the notion of intensional inclusion (Lambert and van Fraassen, 1970):

> P is intensionally included in Q iff ⌜Necessarily, all the individuals which are P are Q⌝ is true.

and

> *P* is intensionally included in *Q* iff ⌜All (possible) individuals which are *P* are *Q*⌝.

We have kept our semantics free from modalities and independent of modal logics for several reasons. Firstly because we do not need them. Secondly because it is far from clear how modal prefixes should be interpreted. If all 'necessarily' is intended to mean is logical necessity, then the concepts of entailment do this job and much better: they elucidate a notion of relative (not absolute) necessity – the necessity of a conclusion relative to its premises and to the accepted rules of inference. And if what is meant is ontic (or physical) necessity, then it has no place in semantics – nor, for that matter, does modal logic supply an adequate elucidation of that concept. In sum, I find no use for modal logic in semantics. On the other hand semantics should be used to try and solve some of the riddles of modal logic, such as whether or not the statement ⌜That trick may work⌝ means the same as ⌜That trick may not work⌝.

So far we have looked into one of the dimensions of sense. We now turn to the remaining dimensions – purport and import.

CHAPTER 5

GIST AND CONTENT

Our analysis of sense in the previous chapter has been "local" or "horizontal": it was limited to the given construct with little regard for its logical relatives. The limitations of such an approach are obvious. For one thing it cannot do justice to propositions of the form $\ulcorner A$ means $B\urcorner$, where 'means' stands for "entails" or for "is entailed by". For example, that x is loved "means" (follows from) somebody loving x; it also "means" (entails) that x is lovable. In the present chapter we shall supplement the "horizontal" or "local" approach to sense with a "vertical" or "global" analysis. We shall in fact ask ourselves the questions: What is the ancestry (or set of implicants) of a construct?, and What is the progeny (or set of implicates) of a construct? In other words, we will elucidate the notions we have called *purport* and *import*.

We shall formalize the ideas (*a*) that the purport of a construct in a given context is the collection of constructs upon which it depends, or which determine it (logically), and (*b*) that the import of a construct in a given context is the collection of constructs that hang from it, or that are determined by it (logically). The concepts of purport and import are thus mutually dual. And both are context-dependent: the purport and the import of a construct depend on the body of knowledge in which it occurs. This relativization of sense to context curtails a certain freedom – the freedom to make arbitrary meaning shifts. The advantage of such a relativization is clear: it is well-nigh impossible to determine the exact sense of a stray predicate – whence the endless semantic disputes in the young growing fields as well as in the old undisciplined areas. Only systemic predicates and statements, i.e. constructs belonging to definite deductive systems, have definite purports and imports. (When transplanted to a different theory, if not rejected, a construct may well acquire a new sense – i.e. it may become a somewhat new construct.)

Finally the full sense of a construct in a given context may be taken to equal the union of its purport and its import in that context. Consequently the intension or "inner sense" of the construct will be included in its full sense. But, since enjoying a definite place in an axiomatic pecking order is the exception rather than the rule, we shall be well advised to keep the theory of intensions even if intensions are only parts of full senses.

1. CLOSED CONTEXTS

1.1. *Closed Contexts and Their Structure*

Recall the notion of a context introduced in Ch. 2, Sec. 3.4. A context $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ is a construct composed of a set S of statements, the extralogical predicates occurring in which belong to a family $\mathbb{P}$ of predicates all of which refer to individuals in a domain D of objects.

A context may be amorphous or structured. See Table 5.1.

TABLE 5.1

Conceptual systems

Item	Structure	Semantics
Set of constructs	———	———
Context	———	Referential homogeneity
Closed context	Boolean algebra	Referential homogeneity
Theory	Filter	Referential homogeneity
Consistent theory	Ultrafilter	Referential homogeneity

We get a structured context if we keep D and $\mathbb{P}$ fixed and allow all and only logical operations in the sets $\mathbb{P}$ and S. Such a context will be closed both formally (syntactically) and as regards reference (semantically). The former because logical processing will produce nothing outside S, the latter because no referents foreign to D will be allowed to intervene in the course of that processing. (Such a double closure need not stifle research: we are always at liberty to leap to a different context.) More explicitly, we lay down

DEFINITION 5.1 The structure $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ is called a *closed context* iff (i) $\mathbb{C}$ is a context and (ii) S is closed under negation, conjunction, disjunction, and generalization (both existential and universal).

From an algebraic point of view the statements in a closed context constitute a complemented lattice. Indeed every member s of S has its opposite number $\neg s$ in S and, for any two elements s and t of S, both the join $s \vee t$ and the meet $s \wedge t$ (i.e. the conjunction $s \,\&\, t$) are in S. Moreover $\vee$ distributes over $\wedge$ and conversely, so that the lattice is distributive in addition to being complemented. And because of the latter property S contains a null element □ as well as a universal element ▯. Every extensionally void statement equals □ and every extensionally universal statement equals ▯. In brief, we have

THEOREM 5.1 The statements in a closed context constitute a distributive complemented lattice with zero element and unit element – i.e. a Boolean algebra.

A similar result holds for certain subsets of the collection of predicates in a closed context. To prove this we recall how the basic connectives are defined for predicates (Ch. 1, Sec. 2.2).

DEFINITION 5.2 Let $\mathbb{C} = \langle S, \mathbb{P}, D \rangle$ be a closed context. If P and Q belong to $\mathbb{P}$, and are both defined on a subset $E \subseteq D$, then

(i) $\neg P : E \rightarrow S$, with $\neg Px = \neg (Px)$ for every x in E;

(ii) $P \wedge Q : E \rightarrow S$, with $(P \wedge Q)\, x = Px \wedge Qx$ for every x in E.

Consider now the totality of predicates in $\mathbb{P}$ with domain $E \subseteq D$, i.e. $\mathbb{P}_E = S^E$. Since by hypothesis S is a lattice and moreover a Boolean algebra, mathematics tells us that $\mathbb{P}_E = S^E$ has the same algebraic structure. In other words, we have

THEOREM 5.2 The totality of predicates defined on the same domain and belonging to a closed context form a complemented distributive lattice with zero element and unit element – i.e. a Boolean algebra.

By virtue of the algebraic similarity between the set S of statements and the subset $\mathbb{P}$ of predicates in a closed context, we may call the two by the generic name *closed set of constructs*. In this way Theorems 1 and 2 can be lumped into the

PROPOSITION Every closed set of constructs is a Boolean algebra.

This formal unification of statements and predicates will save us ink provided we remember that it is not the whole set $\mathbb{P}$ of predicates in a closed context, but only those subsets of $\mathbb{P}$ constituted by predicates with equal domains, that have the said Boolean structure. For example, "wave length" and "refractive index" belong to wave optics, which is a theory and therefore a closed context, but since they are not defined on a common domain they do not belong to a closed predicate context.

1.2. *The Logical Ancestry of a Construct*

Since the purport of a construct in a context will prove to be the principal ideal generated by the construct, it will pay to recall the ABC of ideals. An ideal in a lattice L is that subset I of L that contains all the predecessors of any given element of L as well as all the joins of any two elements of I (including of course the join of a member with its complement). More explicitly we have the following definition:

If L is a lattice, then I is an *ideal* in L iff I is a nonempty subset of L satisfying the following conditions:

*I*1 For all $x \in I$ and $y \in L$, if $y \leqslant x$ then $y \in I$.

*I*2 For all $x, y \in I$, $x \vee y \in I$.

Clearly every lattice is an ideal. (Even the singleton composed of the null construct is an ideal.) In particular, our closed sets of constructs are ideals. And removing the universal element $\Box$ from I turns I into a *proper ideal* $I \subset L$. Finally, making sure that $\Box$ is in a proper ideal authorizes us to pronounce it a *maximal ideal.* In other words, a proper ideal is a maximal ideal iff, for every $x \in L$, both x and its complement $\bar{x}$ are in I.

Let us apply these algebraic concepts to our lattices of constructs in closed contexts. Such an application requires only one more step, namely to identify the order relation. If the basic set happens to be one of either predicates or statements, then the order relation is that of entailment $\vdash$. That is, if x and y are in a closed set $\mathbb{C}$ of constructs, then

$$x \leqslant y \quad \text{iff} \quad x \vdash y.$$

It is now easy to check that every closed set of constructs from which the universal construct is missing, is a maximal ideal. For reference:

THEOREM 5.3 Let $\mathbb{C}$ be a closed set of constructs partially ordered by the relation $\vdash$ of entailment. Then $\langle \mathbb{C}, \wedge, \vee, \neg, \Box \rangle$ is a maximal ideal.

Every ideal can be analyzed into as many partial ideals (or subideals) as elements in the base lattice. Indeed, every element in the lattice drags the whole lot of its predecessors to form its own private ideal or family tree. More precisely, we have the following definition:

For every element x in a lattice L, the set $\{y \in L \mid y \leqslant x\}$ is called the *principal ideal* generated by x in L. Symbol: $(x)_L$.

The adaptation of these algebraic concepts to our needs yields the following

DEFINITION 5.3 Let x be in a closed set of constructs $\mathbb{C}$. Then the *logical ancestry* of x in $\mathbb{C}$ is the principal ideal generated by x in $\mathbb{C}$:

$$(x)_{\mathbb{C}} = \{y \in \mathbb{C} \mid y \vdash x\}.$$

Note that the ancestry of a construct depends on the context in which it occurs.

We proceed to put the preceding ideas to work in semantics.

2. SENSE AS PURPORT OR LOGICAL ANCESTRY

2.1. *Purport and Gist*

The purport of a construct, we said in Sec. 1, may be regarded as the totality of its conceptual determiners, i.e. the collection of constructs that "define" or prove it. A natural explication of this intuitive idea is that of logical ancestry introduced by Definition 3. More explicitly we stipulate

DEFINITION Let x be a construct belonging to a closed set of constructs $\mathbb{C}$. Then the *purport* of x in $\mathbb{C}$ equals the logical ancestry of x in $\mathbb{C}$, i.e. the principal ideal generated by x in $\mathbb{C}$:

$$\mathscr{P}\!ur_{\mathbb{C}}(x) = (x)_{\mathbb{C}} = \{y \in \mathbb{C} \mid y \vdash x\}.$$

For example, because different proofs of a given theorem may exhibit different connections, the purport of a theorem will depend on the premises employed to prove it. This change in purport is corresponsible for the diversity in pragmatic meanings, i.e. for the fact that the same theorem may mean different things to different people (cf. Wang, 1966).

Of all the determiners of any given construct x some may be more basic than others and, by the same token, they may point to whatever is essential in x. In other words, the set of ultimate determiners sufficient to determine x constitute the basic purport or gist of x. More explicitly, we make

DEFINITION 5.5 If x is a member of a closed set of constructs $\mathbb{C}$, then the *gist* (or *basic purport*) of x in $\mathbb{C}$ is the smallest subset of elements of $\mathbb{C}$ that entail x.

For example, the gist of "$\emptyset$" in the algebra of sets consists of two statements: one to the effect that the union of $\emptyset$ with an arbitrary set equals the latter, and the other to the effect that the intersection of $\emptyset$ with an arbitrary set equals the former. We shall return to the concept of gist in the next subsection.

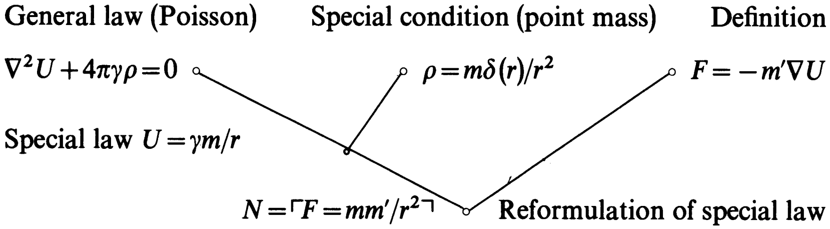
If an ideal containing the null element $\square$ is deprived of the unit element ▯, it turns into a maximal ideal. (Recall Sec. 1.2.) Now, removing the universal construct ▯ from a closed set of constructs amounts to keeping only the nontautologous constructs, i.e. the extralogical ones. This justifies

DEFINITION 5.6 Let x belong to a closed set of constructs $\mathbb{C}$. Then the *extralogical purport* of x in $\mathbb{C}$ is the maximal principal ideal generated by x in $\mathbb{C}$, i.e.

$$\mathscr{P}ur_{\mathbb{C}\bar{L}}(x)=\{y\in\mathbb{C}\mid y\vdash x \quad \text{and} \quad y\neq\text{▯}\}$$

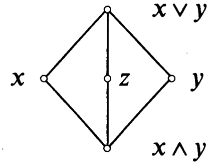
Remark 1 The extralogical purport of a logical item in $\mathbb{C}=L$ is nil. *Remark 2* If $\mathbb{C}$ is a nonlogical theory (in the sense that it includes not only logical predicates but also extralogical ones), then the extralogical purport of any construct in $\mathbb{C}$ consists of a set of mathematical or factual items. *Remark 3* It is not that tautologies are taken out of the way of deduction in order to find the extralogical purport of some construct: if this were done then deduction would hardly be possible. Tautologies are removed once the deductive process is completed. They are counted as midwives or catalyzers rather than as ancestors of the constructs we are interested in.

Example Consider the most famous, though not the best understood, of all law statements: Newton's law of gravitation $N=\ulcorner F=\gamma mm'/r^2\urcorner$. In the context of Newton's own theory of gravity the law is an axiom and thus constitutes its own purport. But in a wider context, such as the classical theory of gravitation, N is derived from a far more general law (Poisson's) as a special case, namely when the source mass m is concentrated at a point. In this case the mass density can be set equal to $m\delta(r)/r^2$, where δ is the Dirac delta. The skeleton of the deductive tree (i.e. if all the mathematical auxiliaries are discarded) is this:


General law (Poisson)
Special condition (point mass)
Definition
$\nabla^2 U + 4\pi\gamma\rho = 0$
$\rho = m\delta(r)/r^2$
$F = -m'\nabla U$
Special law $U = \gamma m/r$
$N = \ulcorner F = mm'/r^2 \urcorner$
Reformulation of special law

Purport of $N = \{\nabla^2 U + 4\pi\gamma\rho = 0,\ \rho = m\delta(r)/r^2,\ F =_{\mathrm{df}} -m'\nabla U,\ U = \gamma m/r,\ N\}$ (Actually this, when "read in factual terms", i.e. upon being enriched with the requisite semantic assumptions, is the *factual* purport of Newton's law of gravity: see Sec. 2.3.)

The next desirable step in the development of our theory would be to set up a calculus of purports enabling one to compute, say, $\mathscr{P}ur_{\mathbb{C}}(x \wedge y)$ out of $\mathscr{P}ur_{\mathbb{C}}(x)$ and $\mathscr{P}ur_{\mathbb{C}}(y)$. Yet, it is not apparent how this could be done: there are no obviously regular or lawful relations among the principal ideals in a given lattice. In particular, it is not the case that, for arbitrary elements x, y in a general lattice L, $(x \vee y)$ will equal either $(x) \cup (y)$ or $(x) \cap (y)$. A simple counterexample is this:


$x \vee y$
x
z
y
$x \wedge y$

$$(x \vee y) = (x) \cup (y) \cup (z)$$

The only apparent simple relations between purports are shown in

THEOREM 4 For any two constructs x, y in a set of closed constructs $\mathbb{C}$,
(i) $\mathscr{P}ur_{\mathbb{C}}(x) = \mathscr{P}ur_{\mathbb{C}}(y)$ iff $x = y$
(ii) If $x \vdash y$ then $\mathscr{P}ur_{\mathbb{C}}(x) \subseteq \mathscr{P}ur_{\mathbb{C}}(y)$.

2.2. *The Gist of a Basic Construct*

Let us now apply the foregoing considerations to the most important constructs in a well organized theory, namely its basic (primitive) concepts and its basic assumptions (axioms). Since such constructs have no logical ancestry other than themselves, they are their own purports and moreover their own gists. More precisely, we have the following

COROLLARY 5.1 In an axiomatized theory the gist of an axiom is the singleton constituted by the axiom itself, and the gist of a primitive concept is the set of axioms in which it occurs.

Example 1 In classical particle mechanics (as axiomatized in Bunge, 1967b) the gist of the primitive concept M of mass is determined jointly by the following postulates:

Mathematical axiom M is an additive function from the set P of particles to the non-negative real numbers.

Factual axiom Newton's second law of motion.

Semantic axiom $M(p)$ represents the inertia of particle $p \in P$.

Example 2 The gist of the concept $\rhd$ of biological ancestry, a primitive in the theory of evolution as axiomatized by Williams (1970), could be given by the following postulates:

Mathematical axiom $\rhd$ is a strict partial ordering of the set B [of organisms].

Factual axiom For every x in B: $\neg(x \rhd x)$.

Semantic axiom $\rhd \mathrel{\hat{=}}$ ancestry relation. [I.e. if x and y are distinct individuals in the set B of organisms, then '$x \rhd y$' designates $\ulcorner x$ is an ancestor of $y\urcorner$.]

Finally an important qualification. Since independent axioms (i.e. not interdeducible postulates) are separate objects, it makes sense to speak of their separate gists – i.e. of themselves. In some cases it is also meaningful to speak of the gists of distinct basic *concepts*, or primitives. For example, in a meet semilattice $\langle S, \leqslant, \wedge \rangle$ the gist of $\leqslant$ does not overlap with that of $\wedge$, as these two concepts are not only mutually independent but also subject to separate axioms. However, this situation is exceptional: usually an axiom contains two or more primitive concepts that determine one another, so that it becomes impossible to reckon their individual gists even though the concepts themselves are distinct and not interdefinable. In other words, in general mutually independent concepts fail to have disjoint gists. An example will make this clear.

Suppose we enrich a meet semilattice with the join operation $\vee$ to obtain a lattice $\langle S, \leqslant, \wedge, \vee \rangle$. Now the two operations, meet and join, intertwine in such a way that it is no longer possible to disentangle their separate gists. Better: they have no separate gists. In effect, on top of the separate axioms for $\wedge$ and $\vee$, such as those of associativity, we now have the typical absorption laws

$$x \wedge (x \vee y) = x, \quad x \vee (x \wedge y) = x.$$

We can still *distinguish* the two concepts: meets precede joins. But we cannot *disentangle* their individual senses, for they overlap. In other words, the meet and the join operations have no separate gists. Only the *theory* of $\wedge$ and $\vee$ as a whole, to wit, lattice theory, has a definite gist – namely the set of its postulates.

Because the gist of a construct is included in its full purport, which is in turn part of the full sense of the construct, we conclude that *the unit of sense is the whole body of theories in which the construct occurs.* A single statement never suffices and a single theory is often insufficient because a concept may do different things at different places. Unless this point is kept in mind, our thesis that purport is included in sense may be criticized by exhibiting alleged consequences that are counterintuitive (Castonguay, 1972). Thus take a biological theory B containing the assumption that all organisms contain carbon. If one forgets where "carbon" comes from, namely chemistry, one will set $\mathscr{P}ur$ ("carbon") = {"organism"}, which is obviously inadequate. The force of this argument against purport consists in its very weakness – i.e. that it ignores that the concept of carbon is not fully characterized by the biological theory B, which borrows it from chemistry. It is the latter that determines the gist (though not the full sense) of "carbon" as consisting of a set of statements containing *inter alia* the definite description "The element with atomic number 6". Our view is that the full sense of "carbon" is determined by the whole lot of theories containing this concept. B is one of them and it contributes just one of the genuine ingredients of the sense of "carbon", namely the statement that carbon enters the composition of all organisms. In conclusion, a meaning analysis of an isolated construct is bound to yield only a partial, hence misleading result.

Let us now take a closer look at the gist of theories.

2.3. *The Gist of a Theory*

We shall now apply the Definition 5 of gist to a theory, or hypothetico-deductive system. This is a very special kind of closed set of constructs: it is a set of statements glued by the relation of deducibility. (In algebraic terms to be clarified in the next section: A theory is a certain subset of a Boolean algebra of statements, namely that which does not contain the null element but contains all the followers of any given element as well as the meet of any two elements. A theory is, in short, a proper filter on a Boolean algebra of statements. And a consistent theory is an ultrafilter, i.e. a proper filter such that, for any element in the Boolean algebra, either the element or its complement is in the filter but not both.)

To determine the gist of a theory it is best to axiomatize the latter. Indeed, the process of tracing the logical ancestry of any formula in the theory leads to and terminates in some or all of the initial statements of the theory. Among such initial formulas we count not just the specific axioms of the theory but also all of its basic presuppositions, i.e. all the axioms belonging to different contexts and employed in the theory (e.g. as lemmas) without being questioned. The totality of such basic presuppositions may be called the *background* of the theory. For example, ordinary logic and variable portions of mathematics belong to the background of every scientific theory. Many scientific theories presuppose, in addition to such formal theories, other (more fundamental) scientific theories. For example modern solid state theories presuppose quantum mechanics. In sum the search for the gist of a theory must not overlook the background of the theory.

The preceding considerations justify

DEFINITION 5.7 Let T be a theory based on a set A of axioms and on a background B, and call L the set of tautologies included in A or in B. Then

(i) the *gist of* T equals $A \cup B$;
(ii) the *extralogical gist* of T equals $(A \cup B) - L$;
(iii) the *specific extralogical gist* of T equals A.

A theory will have a *factual purport* just in case it includes semantic formulas (in the sense of Ch. 3) specifying what factual items the basic concepts of the theory refer to and which if any traits of the referents those concepts are deemed to represent (whether accurately or not). Such semantic assumptions accompany two other sets of axioms in an axiomatized scientific theory (Bunge, 1967d, 1967f, 1973a, 1973b). One of the sets is constituted by the basic factual formulas of the theory, such as the hypotheses concerning the constitution, structure, and mode of change of the systems concerned. The other set of basic elements is formed by the assumptions concerning the mathematical nature of the concept occurring in the previous statements. In other words, the axiom base A of a scientific theory can be partitioned into the following sets:

M = *The mathematical postulates* (e.g. function differentiability)

F = *The basic factual formulas* (e.g. law statements)

S = *The semantic formulas* (designation and denotation rules and representation assumptions).

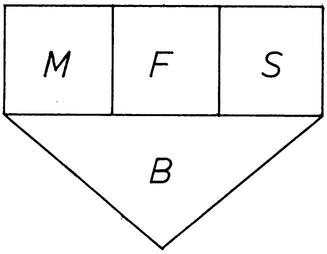


Fig. 5.1. The foundations of a scientific theory: the background B, the mathematical assumptions M, the factual hypotheses F, and the semantic formulas S.

Prima facie F is purely factual whereas M and S are nonfactual. However, a formula has a factual purport provided it is "read in the light" of both the mathematical postulates and the semantic assumptions: otherwise the formula is mathematically as well as factually senseless. Hence it is impossible to single out a subset of A as constituting the factual gist of T. Put in a positive way: All of the nonlogical axioms of a theory constitute its specific factual gist. To repeat with solemnity:

DEFINITION 5.8 Let T be a theory with axiom base $A = M \cup F \cup S$, where M is the set of mathematical axioms, F the specific factual assumptions, and S the semantic formulas of T. Then

(i) the *specific mathematical gist* of T equals $M \cup F$;

(ii) the *specific factual gist* of T equals $M \cup F \cup S$, i.e. the specific extralogical gist of T – in short the whole of A.

This is no concession to semantic holism. We are not claiming that the factual purport of a theory is unanalyzable, but that it is the outcome of combining three distinct reactants: mathematical postulates, factual hypotheses, and semantic formulas. We can distinguish these ingredients but not separate them: M is idle without F, F meaningless without both M and S, and S pointless without F. The three components are no less united and no less distinguishable than the sides of a triangle.

2.4. *Changes in Gist*

The 1960's heard a number of spirited quarrels concerning the changes in meaning accompanying scientific revolutions. Unfortunately there was more noise than light about them, for they were not preceded by an elucidation of what was rightly deemed to change, namely meaning. Let us see whether we can do any better with the help of the tools wrought in Sec. 2.3. (A fuller investigation of the question will have to wait until Ch. 7, Sec. 3.3.)

Our Definitions 7 and 8 provide a simple way of elucidating the concept of change in one of the components of meaning, namely basic purport or gist. Let us see how this works in a simple case before proceeding to a general definition.

The simplest algebraic structure is a set S together with a binary operation $\circ$ in S. This structure, called a *groupoid*, may be characterized by a single axiom;

$$A = \ulcorner \circ \text{ is a binary operation on the set } S. \urcorner$$

Let us now add the condition that $\circ$ be associative. The ensuing structure is called a *semigroup*. The axioms of the theory of semigroups are the previous statement A and

$$B = \ulcorner \circ \text{ is associative in } S. \urcorner$$

Consequently by Definition 7 the specific extralogical gist of the theory of semigroups is $\{A, B\}$. The only difference between this theory and the theory of groupoids is constituted by the additional postulate B. That is

$$\delta\,(\text{Semigroup theory, Groupoid theory}) = \{A, B\} - \{A\} = \{B\}.$$

The generalization suggests itself: The difference in basic purport (gist) between two theories consists of the axioms they do not share. More explicitly we propose

DEFINITION 5.9 Let T and T' be two nonlogical theories with axiom bases A and A' respectively. Then the *difference in specific extralogical gist* between T and T' equals the symmetric difference between A and A':

$$\delta(T, T') = A \Delta A' \equiv A - A' \cup A' - A.$$

It sometimes happens that two scientific theories share a mathematical formalism but interpret it differently: that is, they have the same M's and F's but different S's. At times this semantic difference is due to a difference in the referents, at other times the referents are the same but the predicates are assumed to represent different traits of their referents. An example of the first kind of semantic unlikeness is afforded by the theories of contagion and of the spread of rumours. An example of the second kind of semantic difference is the variety of rival interpretations of quantum mechanics. In either case the difference lies in the semantic formulas of the theory. More exactly we have

DEFINITION 5.10 Let T and T' be two theories with the same mathematical gist but different sets S and S' respectively of semantic formulas. Then the *difference in specific factual gist* between T and T' equals the symmetric difference between S and S':

$$\delta(T, T') = S \Delta S'.$$

From a semantic point of view it does not matter whether T and T' are coexistent or successive. Nor is it indispensable, in order ot use the concept of change in gist, to axiomatize the theories of interest. Axiomatization is needed only to determine the exact amount of change in basic purport or gist.

The problem of changes in both full sense and reference, i.e. the question of meaning change, will be taken up in Ch. 7, Sec. 3.3. We must now turn to the dual of purport, namely import, which is the other component of sense. (Since constructs deriving from the same source have the same purport, they must differ in import if they are to differ at all.)

3. SENSE AS IMPORT OR LOGICAL PROGENY

3.1. *The Logical Progeny of a Construct*

In Sec. 1 we characterized the import of a construct as its logical offspring. Since the concept of a filter will allow us to elucidate the notion of logical progeny, we may as well take a look at filters. We can do it quickly because a filter is the dual of an ideal – a character we met in Sec. 1.2.

In Sec. 1.1 we learned that every closed set of constructs is a Boolean algebra B and, *a fortiori*, a lattice. When tracing down the followers of an element (predicate or statement) in such an algebra we wish to avoid the surprise of ending up in the null element, for this spells disaster just as the unit element spells triviality. To get rid of $\square$ we take a certain subset of B known as a proper filter. But first the general notion.

A filter in a lattice L is that subset F of L which contains all the followers of any given element of L as well as the meets of any two elements of F. More precisely,

If L is a lattice, then F is a *filter* in L iff F is a nonempty subset of L satisfying the following conditions:

$F1$ For all $x \in F$ and $y \in L$, if $x \leqslant y$ then $y \in F$.
$F2$ For all $x, y \in F$, $x \wedge y \in F$.

If the null element is missing from F the filter is called *proper*. Since our construct lattices are Boolean, our filters are more than proper: they are ultrafilters. In other words, a proper filter F in a Boolean algebra B is called an *ultrafilter* just in case, for each $x \in B$, either x or its complement $\bar{x}$ is in F but not both.

Our closed sets of constructs are ultrafilters provided we remove the null construct from them. This will suffice to guarantee that they are consistent. For, if x is in F, then $\bar{x}$ will be automatically excluded from F, as a consequence of which $x \wedge \bar{x}$ will be out as well. In short, we have the following

LEMMA A closed set of constructs is *consistent* iff it is an ultrafilter.

Finally, a filter can be analyzed into partial filters (subfilters). Every member of the base lattice generates its own family, that constitutes the private filter of that element. Formally:

For every element x in a lattice L, the set $\{y \in L \mid x \leqslant y\}$ is called the *principal filter* generated by x. Symbol: $)x(_L$.

Since in the case of constructs the ordering relation is that of entailment, for our purposes we need

DEFINITION 5.11 Let x be in a closed set $\mathbb{C}$ of constructs. Then the *logical offspring* of x in $\mathbb{C}$ is the principal filter generated by x in $\mathbb{C}$:

$$)x(_{\mathbb{C}}=\{y\in\mathbb{C} \mid x\vdash y\}.$$

The offspring or progeny of a nontautologous construct depends on the context in which it occurs. (Tautologies, being deductively barren, have no progeny.)

It should come as no surprise the manner in which these concepts, taken from the algebra of logic, will presently be exploited for semantic purposes.

3.2. *Import*

We are now in a position to elucidate the idea that the import of a construct is the collection of constructs it generates or "contains":

DEFINITION 5.12 The *import* of a construct x in a closed set of constructs $\mathbb{C}$ equals the former's logical offspring, i.e. the principal filter it generates:

$$\mathscr{I}\!m\!p_{\mathbb{C}}(x)=)x(_{\mathbb{C}}=\{y\in\mathbb{C} \mid x\vdash y\}.$$

Example 1 The import of a tautology is nil. *Example 2* Take identity to be defined by Leibniz' law. Then the import of "=" includes the reflexivity, symmetry and transitivity of "=". (This would have puzzled Austin, who regarded "same" as being devoid of positive meaning. But this is excusable: neither Wittgenstein nor his followers had a semantic theory of meaning.) *Example 3* The import of the law of the vibrating string includes the superposition principle (⌜The superposition of any two oscillations is a third oscillation.⌝)

If $\mathbb{C}$ happens to be a *consistent* closed set of constructs, i.e. one in which no contradictions occur, then the import of a construct x in $\mathbb{C}$ will be the principal *ultra*filter generated by x in $\mathbb{C}$. Because filters are the duals of ideals, and ultrafilters the duals of maximal ideals, we see that import and purport are the dual of one another, hence mutually complementary rather than rivals. (More on this in Sec. 4.)

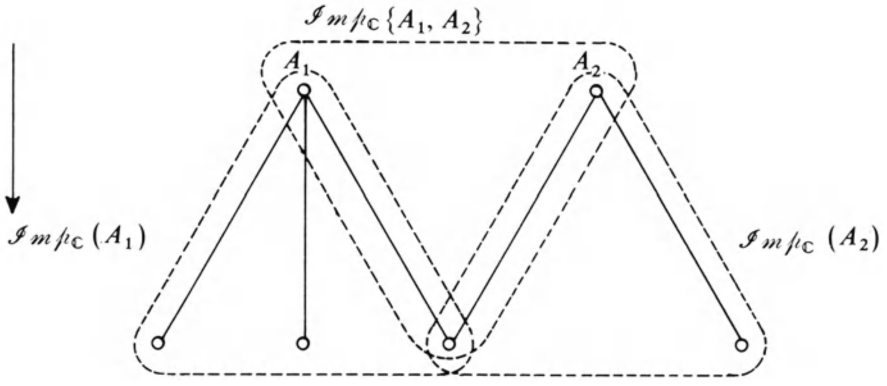
It would be pleasant, but it seems chimerical, to obtain general relations among the principal filters generated by arbitrary constructs in a closed set of constructs, so as to compute, say, the import of the conjunction of two statements in terms of their individual imports. A similar obstacle was met in Sec. 2.1 with reference to principal ideals. A definite algebraic structure is found only in each offspring of a given construct as well as in the total context. The only simple statement we can make about the relations in import between two constructs is the dual of Theorem 4, namely

THEOREM 5.5 For any two constructs x, y in a closed set of constructs $\mathbb{C}$,
(i) $\mathcal{I}mp_{\mathbb{C}}(x)=\mathcal{I}mp_{\mathbb{C}}(y)$ iff $x=y$;
(ii) If $x \vdash y$ then $\mathcal{I}mp_{\mathbb{C}}(x) \supseteq \mathcal{I}mp_{\mathbb{C}}(y)$

On the other hand it is possible and advantageous to extend the concept of import to both a set of constructs and their conjunctions. Let $\mathbb{C}_0$ be a subset of a closed set of constructs $\mathbb{C}$. Usually there will be a number of filters containing $\mathbb{C}_0$. And the intersection of any nonempty set of filters is a filter provided it is nonempty. It is, of course, the smallest filter containing $\mathbb{C}_0$ and for this reason it is called the *least filter generated* by $\mathbb{C}_0$. Accordingly we stipulate

DEFINITION 5.13 Let $\mathbb{C}_0$ be a subset of a closed set of constructs $\mathbb{C}$. Then the *minimal import* of $\mathbb{C}_0$ equals the least filter generated by $\mathbb{C}_0$ in $\mathbb{C}$.

Example Let $\mathbb{C}$ be a theory and $\mathbb{C}_0$ some of its axioms. Then $\mathcal{I}mp_{\mathbb{C}}(\mathbb{C}_0)$ is the smallest collection of theorems (including $\mathbb{C}_0$) entailed by $\mathbb{C}_0$. Look at the picture:


$\mathcal{I}mp_{\mathbb{C}}\{A_1, A_2\}$
A_1
A_2
$\mathcal{I}mp_{\mathbb{C}}(A_1)$
$\mathcal{I}mp_{\mathbb{C}}(A_2)$

We shall get a far more interesting semantic concept if we consider the offspring of a bunch of axioms taken jointly. Let us look into this.

3.3. *Theory Content*

Consider a scientific theory T based on a set $\{A_1, A_2, \ldots, A_m\}$ of specific axioms as well as on a background $\{B_1, B_2, \ldots, B_n\}$ of presuppositions. The latter include, among others, all the logical and mathematical theories underlying T. Let us conjoin all such initial assumptions and examine their import. Call

$$A = A_1 \,\&\, A_2 \,\&\ldots\&\, A_m,$$

The import of $A \,\&\, B$ is, of course, the offspring of $A \,\&\, B$. One part of this import is purely logical: it consists of the set L of tautologies included in B or entailed by $A \,\&\, B$. If we remove this part L we are left with the extralogical import, or *content*, of the axiom basis $A \,\&\, B$. More explicitly, we make $B = B_1 \,\&\, B_2 \,\&\ldots\&\, B_n$.

DEFINITION 5.14 Let T be a theory based on a conjunction A of axioms and another conjunction B of presuppositions, and call L the set of tautologies in B or in the offspring of $A \,\&\, B$. Then

(i) the *import* of the axiom base $A \,\&\, B$ equals the principal filter generated by $A \,\&\, B$, i.e.

$$\mathscr{I}\!mp(A \,\&\, B) =$$

(ii) the *content*, or *extralogical import*, of the axiom base $A \,\&\, B$ equals the complement of the import of $A \,\&\, B$ relative to L:

$$\mathscr{C}(A \,\&\, B) = \mathscr{I}\!mp(A \,\&\, B) - L.$$

Going into some details of the composition of a scientific theory will allow us to grasp the elusive concept of theory content. As may be recalled from Sec. 2.3, the axiom base of a theory in factual science consists of (a) purely mathematical axioms M, (b) specific factual premises (such as law statements and subsidiary hypotheses) F couched in the terms characterized by M, and (c) semantic formulas S. Is it possible to isolate the factual content of these basic assumptions from their mathematical ingredients? No, because the factual components cannot even be stated, let alone processed and interpreted, in isolation from the other two. What can be neatly detached is the mathematical content of the initial assumptions: this is done by simply removing the semantic axioms S and finding the progeny of the remaining premises. All this is encapsulated in

DEFINITION 5.15 Let T be a theory with background B, mathematical assumptions M, factual assumptions F, and semantic formulas S, where B, M, F and S are taken jointly. Then

(i) the *mathematical content* of the axiom base A & B, or *formalism* of the theory T, equals the offspring of B & M & F with exclusion of L, i.e.

$$\mathscr{C}_M(A \,\&\, B) = \mathscr{I}\mathscr{m}\mathscr{p}(B \,\&\, M \,\&\, F) - L;$$

(ii) the *factual import* of the axiom base, or *factual content* of the theory T, equals the offspring of B & M & F & S with exclusion of L, i.e.

$$\mathscr{C}_F(A \,\&\, B) = \mathscr{I}\mathscr{m}\mathscr{p}(B \,\&\, M \,\&\, F \,\&\, S) - L.$$

That is, the factual content of a scientific theory coincides with the factual import of its axiom base, which is in turn identical to its extralogical import. In other words, while scientific theories have detachable, hence portable mathematical formalisms, their factual content is not detachable from the formalism. (*A fortiori* they have no purely empirical or observational content.) The extralogical import of the axiom base of a theory in factual science is both mathematical and factual.

Note that the notion of content, though dependent upon the metalogical concept of deducibility, is independent of the semantic concept of truth of fact. Hence a theory with a rich content may well prove to be false whereas a more modest theory may turn out to be approximately true. Consequently, *pace* Popper (1966), the degree of truth of a theory cannot be measured by its content. If it were so measurable then no actual empirical tests for truth would be called for: a semantic analysis of a theory would determine its truth value.

We close by noting a possible application of the foregoing ideas to model theory. Intuitively the richer an abstract theory ("formalized language") the fewer models it is bound to have. Thus there are less examples of a semigroup with unit element than of a semigroup. This suggests introducing a comparative concept of theory versatility, or its dual, rigidity, namely thus:

DEFINITION 5.16 Let T and T' be two abstract theories with comparable mathematical contents $\mathscr{C}(T)$ and $\mathscr{C}(T')$ respectively. Then T is *more versatile* (or *less rigid*) than T' iff $\mathscr{C}(T) \subseteq \mathscr{C}(T')$.

3.4. *Empirical and Factual Content*

The various philosophies of science may be regarded as so many views on the content of scientific theories – i.e. as so many semantics of science. Here goes, in telegraphic fashion, a sample of the most interesting among those views:

(1) *Conventionalism.* Scientific theories are handy tools for the processing of experience (or the systematization of appearances). They have neither a factual nor an empirical content: $\mathscr{C}_F(T) = \mathscr{C}_E(T) = \emptyset$.

(2) *Positivism.* Scientific theories are economic summaries and articulations of data. They have an empirical (observational, operational) content only: $\mathscr{C}(T) = \mathscr{C}_E(T)$.

(3) *Mellowed positivism (standard view).* Scientific theories are systematizations of experience propped up by heuristic devices, i.e. the so-called theoretical terms, which have no representational function. Part of the content (our "import") of a statement s in a scientific theory is observational (empirical, phenomenal) provided s is linked to observational items via correspondence rules: $\mathscr{C}_E(s) \subseteq \mathscr{C}(s)$.

(4) *Empirical realism.* Theoretical constructs may refer to unobservables even though their meaning is given by the connections established within the theory between the constructs and observational statements built with the observational vocabulary of the theory. Every scientific statement s has both an empirical content and a factual one (its "surplus meaning"): $\mathscr{C}(s) = \mathscr{C}_E(s) \cup \mathscr{C}_F(s)$.

Of these four views the closest to our own is the last, according to which scientific hypotheses have "a surplus meaning over against their evidential basis" (Feigl, 1950). (For elaborations see Rozeboom (1962, 1970). For criticisms see Hempel (1950) and Nagel (1950).) Although timid realism is preferable to the previous forms of empiricism, it is not completely satisfactory: keeping the myth of the observational vocabulary as belonging to every theory and as the source of meaning of the theoretical vocabulary does violence to scientific theories, all the concepts of which are theoretical. Furthermore it has the fatal consequence that it becomes impossible to ascertain what exactly the theory is about: reference becomes indeterminate (Rozeboom, 1962, 1970). Rather than tolerate such reference uncertainty, as Rozeboom advocates, we regard it as sufficient for disqualifying the theory: reference indeterminacy makes not only for meaning indeterminacy but renders empirical tests impossible. We must either retreat to undiluted empiricism or go realists all the way. Since the former has proved a failure we shall try the latter.

We claim that no scientific theory could possibly embrace its own evidence, and that only theories concerning sentient organisms may be said to have an empirical content. The reasons for denying that the evidence relevant to a theory is part of its content have been given in detail elsewhere (Bunge, 1967a, 1969, 1973b). They will be summarized here. Firstly, the computing of a prediction with the help of a theory requires additional theoretical assumptions (e.g. on the particular composition or structure of the referent) as well as empirical data – and neither can be found in a theory unless it happens to be an *ad hoc* systematization of data. (If the special assumptions were already included in the theory then the latter would be inapplicable to further problems. And if the data were part of the theory then observations would be redundant.) Secondly, the production of empirical data that could serve as evidence for or against a theory calls for the cooperation of auxiliary theories employed in the design and interpretation of the empirical operations – i.e. the evidence could not have been predicted on the strength of the substantive theory alone. Should a scientific theory have an empirical content, then this content would not consist in any evidence but in statements concerning some sector of human experience, such as learning or working, considered as an external item. (More on this in a while.) These are some of the reasons for adopting a realist semantics.

Of the three other views sketched above, conventionalism is untenable if only because it renders empirical testing, as well as the quest for truer and deeper theories, utterly pointless. The second view, undiluted positivism, is still popular in scientific circles but it has been discredited by philosophers long ago, so it need not detain us. (Which does not mean disparaging catechisms for the conversion of the infidel.) Only the third view, which we have dubbed *mellowed positivism*, is still alive and influential though on the decline. It is also the only view that has been expounded and examined in detail (Carnap, 1956, 1963a, 1963b, 1966; Hempel, 1958; Braithwaite, 1959, 1962; Wojcicki, 1966; Przelecki, 1969;

etc.). The way this doctrine handles the concept of empirical or observational content is roughly as follows.

The standard view assumes, first of all, that every scientific theory contains not only theoretical terms, such as 'temperature', but also observational or phenomenal ones such as 'hot'. Second, only the theoretical terms are assumed to be subject to the theoretical statements Θ of the theory. They are interpreted, albeit only in part, by reference to the observational terms: this interpretation is performed by the so called "correspondence rules" C of the theory. Third, the content of a statement is, roughly, what we have called its import. And the observational (or empirical) content of a statement s is defined as the class of all nonlogically true observational statements implied by s – i.e. the set of all those consequences of s that contain observational terms but no theoretical ones. In particular, the observational content of s relative to a theory composed of theoretical statements Θ and correspondence rules C (both sets taken conjunctively), is the observational content of $s \& \Theta \& C$. (For a brief and lucid exposé see Carnap (1963b). For details cf. Suppe (1971). For the alleged definability of theoretical terms as functions of empirical ones: Carnap (1961, 1966).)

The root trouble with this view is, of course, that it rests on the unquestioned assumption that every scientific theory does contain observational or phenomenal terms such as 'hot', 'blue', and 'rough'. In point of fact theories contain only theoretical concepts and these are not interpreted with reference to qualia but to supposedly real things (e.g. genes) and their properties (e.g. the mutation frequency of a given gene). The semantic assumptions in a well constructed theory are not concept-concept affairs in which one of the two partners concerns an experiential item: they are concept-fact correspondences (Ch. 3, Sec. 4.1). Now, if there are no purely observational (hence unanalyzed) concepts in a scientific theory, then there is no observational (or phenomenal or operational or empirical) content left. And since (unlike Feigl) Carnap, Hempel and Braithwaite make no room for factual reference, no content at all is left.

This unintended disaster could have been averted if any real scientific theories had been analyzed in the light of the standard semantics, and if the quest for the Holy Grail of the empirical "basis" of science had been given up in time. If theories are regarded instead as partial and symbolic representations of reality, then a far simpler and realistic semantics of science emerges. Every concept in a theory is pronounced *theoretical* if it is elucidated by the theory or borrowed from the background of the theory. And if the concept has a factual reference then any statement that contains the concept has a factual sense. The facts referred to, or represented by, the theoretical statement are not those the empirical evidence is concerned with. (In fact an evidence statement is restricted to observational items, often under control with the help of instruments describable with the assistance of a number of theories.) No scientific theory is concerned with its own test, much less with the mental states of the experimenter. Finally, the factual content of a statement in a scientific theory is determined by all three basic components of the theory: M, F, and S. (Recall the preceding subsection.)

According to our view, then, all scientific theories have a factual content and none of them embraces its own evidential or empirical "basis". However, a scientific theory *can* have an empirical content – provided it *is* about experience. This is the case with most, perhaps all, theories in the sciences of man. Like all other scientific theories, those in psychology, sociology and history concern facts and so have a factual content (even if inaccurate). But (*a*) these facts are treated as external objects epistemologically on a par with physical facts – which is necessary for objectivity, and (*b*) those theories contain no reference to their own empirical tests – which alone renders tests significant. For example, a model of the decay of empires may refer to feelings and motives – of the imperial subjects not the historian's; and its evidence may consist in documents and material testimonials unearthed with advanced physical techniques – but the model won't be concerned with such means for its own validation or invalidation.

In summary, if s is a theoretical statement then the factual content of s equals its extralogical import. Should s belong in natural science, its empirical content will be nil. But if s belongs in the sciences of man, then s will have an *empirical content* included in its factual content and independent of any evidence relevant to s. No theory has an observational or experiential or phenomenal content: experience is lived through (*erlebt*) not beaten into theory. This view may be dubbed *realist semantics.* It boils down to shifting the focus from experience (always subjective) to fact, and consequently prevents us from falling into subjectivism. It suggests that we stop chasing the ghost of the alleged observational content of theories as well as its attendant ghosts – Ramsey sentences and Craig's reduction.

Now that operationism lies buried we can afford to write a nice epitaph on it. The deceased, though mistaken in the means it employed, was sincere in pursuing a worthy goal – namely the uprooting of inscrutables and the testability of hypotheses. This explains the large following it had and still has in the scientific community. Such a noble cause can now be defended in the following way: Although theoretical statements may lack an empirical import, they should acquire one when conjoined with suitable empirical statements relevant to the former. How this strategy can be implemented may be exemplified as follows.

Consider the statement schema

⌜There is a gravitational field in region *x*.⌝ (1)

belonging to some theory of gravitation. Although fields defy direct observation, according to operationism (1) *means* that, if somebody places a test body *y* in region *x*, he will observe *y* to move (cf. Carnap, 1936). But we know this to be a gross misinterpretation of (1), which refers only to a field, not to a test body, let alone to an observer. However, our construal of the notion of sense as import suggests the following repair: conjoin (1) with

⌜A body acted on by a gravitational field is accelerated & *y* is an observable body & *y* is at *x* [whether or not it was put there by somebody] & *z* is suitably equipped to observe *y* & *z* observes *y*.⌝ (2)

This new statement schema has one theoretical component, viz., the first conjunct, which is a law statement. The remaining components of (2) are empirical and moreover relevant to both the theoretical component and the original formula (1). As a consequence of this empirical ingredient, the conjunction (1) & (2) will entail observational statements and consequently will have an empirical import. Part of the latter will be

⌜*z* sees *y* moving⌝, (3)

which is in fact entailed by (1) & (2). In short, although the "operationally meaningful" statement (3) is not in the import of the original theoretical statement (1), it is in the import of the conjunction of the latter with further items, some of which are empirical. What holds for a single statement is true, *a fortiori*, of entire theories. That is, any substantive scientific theory can be enriched with further statements, at least some of them empirical, to yield observational consequences and thus get ready for observational tests. Thus the goal of operationism, namely to ensure testability, is saved without tergiversating the nature of scientific theories. *Requiescat in pace.*

3.5. *Changes in Import and Content*

The content of a theory is determined by its axiom base but is never known in its entirety. The content of a theory is fixed, however, once its axioms have been adopted. Moreover the content is infinite because so is the set of logical consequences of the axioms. And if the theory involves continua, then its content is nondenumerably infinite: the axioms generate an uncountable offspring. On the other hand the known part of the total content of a theory is of course finite and keeps growing as long as the theory stays alive – i.e. as more theorems in it are proved and fresh applications of it are made. This is the case not only with new theories but also with old theories that are still going strong, such as classical mechanics. There is growth in both cases although the growth *tempi* and mechanisms are not the same.

The growth of knowledge and its mechanisms, internal and external, is a fascinating subject but it lies outside the scope of semantics: they are the concern of the philosophy, history, psychology, and sociology of science. What the semanticist can do in this respect is to compare the import of alternative axiom bases, whether successive or contemporaneous. He can thus help other scholars by supplying measures of the net differences in import or in content, differences that may be interpreted as accompanying the replacement of one theory by another in the favor of the scientific community. Definition 15 will help us to contrive such a measure of change in scientific theories.

According to Definition 15, the factual content of a scientific theory equals the sum total of its nonlogical formulas. Hence the difference in content between two theories boils down to whatever nonlogical formulas the theories faill to share. More precisely, we introduce

DEFINITION 5.17 Let T and T' be two scientific theories with backgrounds B and B', mathematical assumptions M and M', factual assumptions F and F', and semantic assumptions S and S' respectively. Then the *difference in factual content* between T and T' equals the symmetric difference of their factual contents:

$$\delta(T, T')=\mathscr{I}\!\mathit{mp}(B \,\&\, M \,\&\, F \,\&\, S)\,\Delta\,\mathscr{I}\!\mathit{mp}(B' \,\&\, M' \,\&\, F' \,\&\, S')-L.$$

Since T and T' are infinite sets, their difference may prove infinite as well. On the other hand the gist differences introduced in Sec. 2.4 (Definitions 9 and 10) look more manageable because they concern only the axiom bases. However, this greater simplicity of the difference in gist does not render it more suitable as a measure of the difference in sense. In fact, since a theory can usually be axiomatized in alternative ways, the gists of two different axiomatizations of a given body of knowledge are bound to differ even though the total content of the theory is unaltered by the different axiomatic arrangement. Obviously we need both measures. And none of them is needed for practical purposes, such as rewarding the scientist who manages to produce the richest of two rival theories. The assumptions, definitions and theorems in our semantics serve only the modest purpose of elucidating some interesting metascientific ideas. (We shall come back to meaning change in Ch. 7.)

4. FULL SENSE

In the previous chapter we studied the content or interior of a construct apart from its exterior or extension and regardless of its place in a body of knowledge. The corresponding concept of sense, i.e. intension, is the one that suits constructs which, without being totally stray, belong to no definite system or closed context.

In the present chapter we have focused on systemic constructs and have distinguished two components of their sense – namely their purport and their import. The purport and the import of a construct are certainly not external to it but they are not strictly intrinsic either: they are essentially context-dependent. A closed context $\mathbb{C}$ may be pictured as a tree oriented from top to bottom. An arbitrary node or branching point x of $\mathbb{C}$, i.e. a systemic predicate or proposition, is assigned two basic sense components. One is its purport or upward sense $\mathscr{P}\!\mathit{ur}_{\mathbb{C}}(x)$, which equals the principal ideal generated by x in $\mathbb{C}$. The other component is the import or downward sense $\mathscr{I}mp_{\mathbb{C}}(x)$, equal to the principal filter generated by x in $\mathbb{C}$. See Figure 5.2.

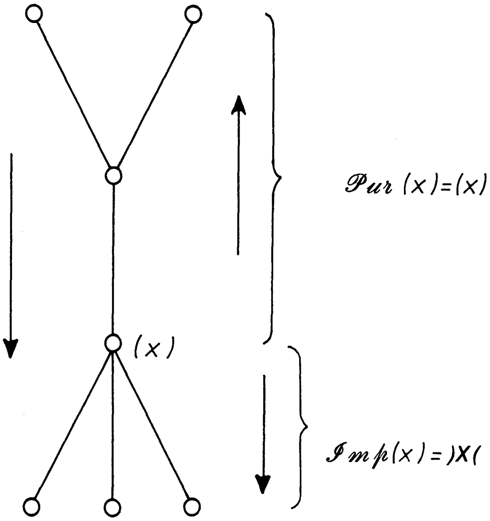


Fig. 5.2. A fragment of a conceptual system. The two components of the sense of a construct x in the system: the purport $\mathscr{P}ur(x)$ and the import

The sense of a construct in a closed context has no components other than its purport and its import. Since both are sets, or may be taken as such, we can build their union. It seems natural then to regard this union as the full sense of the construct of interest. More precisely, we offer

DEFINITION 5.18 Let x be a construct in a closed set of constructs $\mathbb{C}$. Then the *full sense* of x in $\mathbb{C}$ is the union of the purport and the import of x in $\mathbb{C}$:

$$\mathscr{S}_{\mathbb{C}}(x)=\mathscr{P}ur_{\mathbb{C}}(x)\cup\mathscr{I}mp_{\mathbb{C}}(x)=(x)_{\mathbb{C}}\cup)x(_{\mathbb{C}}.$$

This decomposition of the total sense of a construct is consonant with the scientific practice of regarding meanings as being either assumed or derived. Thus if it is *assumed* that the values of a certain function X of time represent positions (or concentrations or population densities), then it can be *inferred* that the value of its time derivative $\dot{X}$ at $t=0$ represents the initial velocity (or rate of change of concentration or population density). For this reason only the factual primitives of a theory need be assigned a factual sense by semantic formulas: the derived (defined or proved) constructs get their meaning from the source, i.e. the axiom base. In other words, in a scientific theory *sense flows downwards*, from assumption to consequence – not upwards, as the positivist doctrine has it.

Since a tautology follows from any proposition and entails all of the other tautologies in a given logic, the full sense of a tautological construct in a context equals the totality of propositions in it plus the underlying logic. Likewise all contradictions are interdeducible, so that the sense of any of them includes all; and since a contradiction entails anything, its sense is the totality of propositions in its context. That is, we have proved

COROLLARY 5.2 Let t be an analytic construct occurring in the logic L underlying a context $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$. Then

(i) $\mathscr{S}_{\mathbb{C}}(t)=S\cup L$;
(ii) $\mathscr{S}_{\mathbb{C}}(\neg t)=S$.

In a strictly logical context, i.e. for $S=L$, the sense of a tautology equals logic itself. In other words, logical truths "say" nothing when confined to logical theories and they "say" too much – in fact everything – when associated with extralogical bodies of knowledge.

Because an arbitrary proposition entails all of the tautologies, it falls between the above extremes of minimal and maximal sense:

COROLLARY 5.3 Let p be a statement occurring in a context $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ with underlying logic L. Then

$$L\subseteq\mathscr{S}_{\mathbb{C}}(p)\subseteq S\cup L.$$

To obtain the extralogical (in particular factual) sense of a proposition we must then subtract L from its full sense:

DEFINITION 5.19 Let p be a proposition in a context $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ with underlying logic L. Then the *extralogical sense* of p in $\mathbb{C}$ is

$$\mathscr{S}_{\mathbb{C}\bar{L}}(p)=\mathscr{S}_{\mathbb{C}}(p)-L.$$

If we know the full senses of two constructs we can ascertain whether they are comparable and, if so, which contains which. Thus the sense of "river" is included in that of "mountain river": the latter is the richer of the two. In general we have

DEFINITION 5.20 Of two constructs with comparable senses, the *richer* (or *more complex*) is the one including the other:
If c and c' are in $\mathbb{C}$, then c is *richer* than $c' =_{\mathrm{df}} \quad \mathcal{S}_{\mathbb{C}}(c) \supseteq \mathcal{S}_{\mathbb{C}}(c')$.

The richer of two concepts is also the more specific and complex, i.e. the less generic and less simple. The last convention serves then as a definition of *semantic complexity* – or its dual, semantic simplicity. But of course even if two constructs are not comparable in sense they may share a grain of sense. This notion of sense overlapping is exactified by

DEFINITION 5.21 The *core sense* of a set $\{c_i \in \mathbb{C} \mid 1 \leqslant i \leqslant n\}$ of constructs equals their shared sense:

$$\mathcal{S}_{core}\{c_i \in \mathbb{C} \mid 1 \leqslant i \leqslant n\} =_{df} \bigcap_{i=1}^{n} S(c_i)$$

Example Each of the different explicata e_i, for $1 \leqslant i \leqslant n$, of a given coarse or intuitive concept c, is supposed to recapture a part of the sense of the latter:

$$\mathcal{S}(c) \supseteq \mathcal{S}_{core}\{e_i \in \mathbb{C} \mid 1 \leqslant i \leqslant n \;\&\; e_i \text{ explicates } c\}.$$

If we focus on the specific contribution a basic construct makes, and brush aside the background of the theory, i.e. if the search for ancestors is stopped at the basic construct a itself, the purport of a will include just a, which is already included in its principal filter or import $\mathcal{I}mp_T(a)$. Hence we have

COROLLARY 5.4 The full specific sense of a basic construct a, characteristic of an axiomatized theory T, equals the import of a, i.e.

$$\mathcal{S}_T(a) = \mathcal{I}mp_T(a).$$

Hence the import of a construct measures its importance. On the other hand the greater the purport of a construct, i.e. the larger its dependence upon other constructs, the smaller its importance.

Because in a closed context $\mathbb{C}$ there is no source of sense other than $\mathbb{C}$ itself, intension has no separate existence in it: it is part of the full sense. More explicitly, we conjecture

PROPOSITION 5.1 The intension of a construct x in a closed set of constructs $\mathbb{C}$ is included in the full sense of x:

$$\text{If } x \in \mathbb{C}, \quad \text{then} \quad \mathcal{I}(x) \subseteq \mathcal{S}_{\mathbb{C}}(x).$$

Example Consider the predicates "triangular" ($\triangle$), "equiangular" (A), and "equilateral" (L), all of them referred to closed plane figures in Euclidean geometry (E). Since $A \neq L$, $\mathcal{I}(\triangle \;\&\; A) \neq \mathcal{I}(\triangle \;\&\; L)$. But, because triangle equiangularity implies triangle equilaterality and conversely, $\mathcal{S}_E(\triangle \;\&\; A) = \mathcal{S}_E(\triangle \;\&\; L)$. Moreover the full sense of either complex predicate includes not just A and L but also the whole lot of derived properties of equilateral triangles, which their intensions do not.

From the preceding it follows that, if a construct happens to be basic (i.e. undefined or assumed) in a given context, then its intension coincides with its full sense and also with its import:

PROPOSITION 5.2 The intension of a basic construct a in a closed set $\mathbb{C}$ of constructs equals the full sense of a:

$$\text{If } a \in \mathbb{C} \text{ is basic in } \mathbb{C}, \quad \text{then} \quad \mathcal{I}(a) = \mathcal{S}_{\mathbb{C}}(a) = \mathcal{I}mp_{\mathbb{C}}(a)$$

A change in the basic principles that determine the sense of a construct may alter the latter even if we keep calling it by the same name. As Hilbert explained to Frege (who disliked axiomatic definitions), "Each axiom contributes something to the definition [of the concept of interest] and, as new axioms are introduced, the concept is accordingly changed" (Hilbert, 1900). In short, meaning is contextual.

This concludes the study of the concepts of sense we started in Ch. 4. It may well happen that our exactifications fail to capture the reader's intuitions. One man's sense is another's nonsense. This is quite in order: no technical concept is supposed to satisfy everybody's presystematic ideas. (Recall the example illustrating Proposition 5.1.) If the proposed elucidations are found unacceptable for technical reasons, then surely better theories can be contrived. But once a theoretical explication is accepted, be it in science or in philosophy, it is one's intuitions that are expected to adjust to the theoretical concept – until a better exactification becomes available. (Cf. Bunge, 1962.)

5. CONCLUSION

There are two main views on sense. One is extensionalism, according to which constructs have an exterior, or extension, but no interior or sense. We have found this, the hollow construct view, unacceptable not only in relation to factual science but also to pure mathematics (Ch. 4, Sec. 1.2). The opposite view regards all constructs other than sets as having a content. But there are many differences of opinion concerning the nature of this content (Ch. 4, Sec. 1.1). These differences are manifested, in particular in the question of the relations between the sense of a whole and the sense of its parts. There are two extreme views on this matter: semantic atomism and semantic holism. Let us take a glimpse at them and locate our own view in this perspective.

Semantic atomism holds that "The sense of the whole is determined by the senses of its parts". It has been the most popular view among illuminist philosophers from Hobbes right down to Montague (1970) and Scott (1970). But no semantic atomist seems to have constructed a suitable theory containing formulas that determine the sense of a construct as a function of the sense of its components. In short, semantic atomism has remained at a programmatic stage. The opposite extreme, semantic holism, maintains that "The sense of every item is determined by the whole of knowledge". Its most eminent proponent in recent times has been Quine (1952). This view too is programmatic and moreover doomed to stay in this stage, because nobody can handle the sum total of human knowledge.

Our own semantic theory takes a *via media* and therefore cannot be as charming as either of the extremes. Indeed, on the one hand we have proposed definite formulas for the intension of a composite construct in terms of the intensions of its components (Ch. 4, Sec. 2). However, this analysis is not a term by term resolution and, moreover, it offers no means for determining the sense of a component. If asked how the latter is to be determined, we offer still another theory, according to which the sense of a construct can be determined provided the construct occupies a definite place in a conceptual system. Hence, to the extent to which our theory provides for the sense analysis of a whole, it implements a modest version of the atomistic program. And since the full sense of a part is shown to depend on its role in a whole, our theory satisfies the holistic injunction not to isolate. In short our semantics constitutes a merger of the best (others would say the worse) of both extremes. If accepted, it becomes understandable why each of the two extremes has its appeal – and why none of them got beyond the program stage.

At first blush our *via media*, by admitting that the sense of a (theoretical) construct is relative to the theory in which it occurs, and is thus liable to become a different construct if inserted in a different theory, would seem to encourage subjectivism. Perish the thought. Firstly, only the sense of a theoretical or systemic construct is determined by the whole theory in which it occurs. Extrasystematic constructs, such as the common observational statements, are not theory-dependent – but then their sense is uncertain. Secondly, it is not that "the meaning of a scientific construct is theory-laden", as the fashionable trend has it, but the other way around: theories (even logical theories) are meaning-laden. The meaning of any component of a theory is relative to the theory, i.e. it depends on the role it plays in the theory, but is not "theory-laden" – whatever this metaphor may mean. Thirdly, every authentic scientific theory, even if false, is objective – ergo so is every factual component of it. It is objective both semantically (by its reference to entities that are supposedly out there) and methodologically – for being publicly scrutable. On the other hand the attempts to defend the objectivity of science by anchoring it to an experiential rock bottom may easily lead back to subjectivism, as witnessed by operationism.

To sum up, we have proposed a contextual but not holistic theory of sense. The contextual nature of sense explains why one and the same symbol can be interpreted differently in different contexts and never fully outside some well delineated conceptual framework. But the matter of interpretation deserves another chapter. As a matter of fact we shall devote it the initial chapter of Part II of this work.

CONTINUED IN VOLUME 2: INTERPRETATION AND TRUTH

BIBLIOGRAPHY

Ajdukiewicz, K. (1935). Syntactic connexion. In S. McCall, Ed., *Polish Logic*. Oxford: Clarendon Press, 1967.

Ajdukiewicz, K. (1967a). Intensional expressions. *Studia Logica* **20**: 63–86.

Anderson, A. R. (1972). An intensional interpretation of truth-values. *Mind* **80**: 348–371.

Angel, R. B. (1970). Relativity and Covariance. Ph. D. thesis, McGill University.

Angel, R. B. (1973). Relativity and Covariance. In Bunge, Ed. 1973b.

Angelelli, I., Ed. (1967). *Gottlob Frege, Kleine Schriften.* Hildesheim: Georg Olms.

Arbib, M. A., Ed. (1968). *Algebraic Theory of Machines, Languages, and Semi-groups.* New York and London: Academic Press.

Aristotle. *De interpretatione.* In *Works*, X, ed. W. D. Ross. Oxford: Clarendon Press 1921.

Aristotle. *Analytica poteriora. Ibid.*

Arnauld, A. and P. Nicole (1662). *La logique, ou l'art de penser.* Paris: Flammarion 1970.

Ayer, J. A. (1959). *Logical Positivism.* Glencoe, Ill.: Free Press.

Bar-Hillel, Y. (1970). *Aspects of Language.* Jerusalem: Magnes Press.

Bar-Hillel, Y. and R. Carnap (1953). Semantic information. *Brit. J. Phil. Sci.* **4**: 147–157.

Bar-Hillel, Y., E. I. J. Poznanski, M. O. Rabin, and A. Robinson, Eds. (1961). *Essays on the Foundations of Mathematics.* Jerusalem: Magnes Press.

Bell, J. L. and A. B. Slomson (1969). *Models and Ultraproducts.* Amsterdam: North-Holland.

Blumenthal, L. M. and K. Menger (1970). *Studies in Geometry.* San Francisco: W. H. Freeman & Co.

Bolzano, B. (1837). *Wissenschaftslehre*, 4 vols. Sulzbach: Seidelsche Buchandlung. Abridged Engl. transl. by R. George, *Theory of Science.* Berkeley and Los Angeles: University of California Press 1972.

Bolzano, B. (1851). *Paradoxes of the Infinite.* Transl. D. A. Steele. London: Routledge & Kegan Paul, 1950.

Bourbaki, N. (1970). *Théorie des ensembles.* Paris: Hermann.

Braithwaite, R. B. (1962). Models in the empirical sciences. In E. Nagel *et al.*, Eds. (1962), pp. 224–231.

Brillouin, L. (1962). *Science and Information Theory.* New York: Academic Press.

Bunge, M. (1956). Survey of the interpretations of quantum mechanics. *Amer. J. Phys.* **24**: 272–286. Reprinted in Bunge (1959b).

Bunge, M. (1959b). *Metascientific Queries.* Springfield, Ill.: Charles C. Thomas Publ.

Bunge, M. (1961a). Laws of physical laws. *Amer. J. Phys.* **29**: 518–529. Repr. in Bunge (1963a).

Bunge, M. (1961b). The weight of simplicity. *Phil. Sci.* **28**: 120–149.

Bunge, M. (1962). *Intuition and Science.* Englewood Cliffs, N. J.: Prentice-Hall.

Bunge, M. (1966). Mach's critique of Newtonian mechanics. *Amer. J. Phys.* **34**: 585–596.

Bunge, M. (1967a). *Scientific Research*, 2 vols. New York: Springer-Verlag.

Bunge, M. (1967b). *Foundations of Physics.* New York: Springer-Verlag.

Bunge, M. (1967c). Analogy in quantum mechanics: from insight to nonsense. *Brit. J. Phil.*

Sci. **18**: 265–286.

Bunge, M. (1967d). Physical axiomatics. *Rev. Mod. Phys.* **39**: 463–474.

Bunge, M. (1967e). A ghost-free axiomatization of quantum mechanics. In M. Bunge, Ed., *Quantum Theory and Reality*, pp. 98–117. New York: Springer-Verlag.

Bunge, M. (1967f). The structure and content of a physical theory. In M. Bunge, Ed., *Delaware Seminar in the Foundations of Physics*, pp. 15–27. New York: Springer-Verlag.

Bunge, M. (1968). Problems and games in the current philosophy of natural science. *Proc. XIVth Intern. Congress of Philosophy* I: 566–574. Wien: Herder.

Bunge, M. (1969). What are physical theories about? *Amer. Philos. Quart. Monograph* **3**: 61–99.

Bunge, M. (1970a). Theory meets experience. In H. Kiefer and M. Munitz, Eds., *Contemporary Philosophic Thought*, **2**: 138–165. Albany, N.Y.: State University of New York Press.

Bunge, M. (1970b). Problems concerning inter-theory relations. In P. Weingartner and G. Zecha, Eds., *Induction, Physics, and Ethics*, pp. 287–325. Dordrecht: D. Reidel Publ. Co.

Bunge, M. (1971a). A mathematical theory of the dimensions and units of physical quantities. In M. Bunge, Ed., *Problems in the Foundations of Physics*, pp. 1–16. New York: Springer-Verlag.

Bunge, M. (1971b). Is scientific metaphysics possible? *J. Phil.* **68**: 507–520.

Bunge, M. (1972a). A program for the semantics of science. *J. Phil. Logic* **1**: 317–328. Reprinted in Bunge, Ed. (1973a).

Bunge, M. (1972b). Metatheory. In *Scientific Thought*. Paris–The Hague: Mouton/Unesco.

Bunge, M. (1973a). *Method, Model and Matter*. Dordrecht: D. Reidel Publ. Co.

Bunge, M. (1973b). *Philosophy of Physics*. Dordrecht: D. Reidel Publ. Co.

Bunge, M. (1973c). On confusing "measure" with "measurement" in the methodology of behavioral science. In Bunge, Ed. (1973b).

Bunge, M. (1973d). Meaning in science. *Proc. XVth World Congress of Philosophy* **2**: 281–286. Sofia.

Bunge, M., Ed. (1973a). *Exact Philosophy: Problems, Goals, and Methods*. Dordrecht: D. Reidel Publ. Co.

Bunge, M., Ed. (1973b). *The Methodological Unity of Science*. Dordrecht: Reidel.

Bunge, M. (1974a). The relations of logic and semantics to ontology. *J. Phil. Logic* **3**: 195–210.

Buridan, J. *Sophisms on Meaning and Truth [Sophismata]*. Transl. T. K. Scott. New York: Appleton-Century-Crofts 1966.

Campbell (1920). *Physics: The Elements*. Cambridge: Cambridge University Press. Repr.: *Foundations of Science*. New York: Dover Publications, Inc. 1957.

Carnap, R. (1936). Testability and meaning. *Phil. Sci.* **3**: 419–471; **4**: 1–40.

Carnap, R. (1939). *Foundations of Logic and Mathematics*. Chicago: University of Chicago Press.

Carnap, R. (1942). *Introduction to Semantics*. Cambridge, Mass.: Harvard University Press.

Carnap, R. (1947). *Meaning and Necessity*. Chicago: University of Chicago Press. Enlarged ed. 1956.

Carnap, R. (1952). Meaning postulates. *Phil. Studies* **3**: 65–73.

Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl and M. Scriven, Eds., *Minnesota Studies in the Philosophy of Science*, I: 38–76. Minneapolis: University of Minnesota Press.

Carnap, R. (1958). *Introduction to Symbolic Logic and its Applications*. New York: Dover

Publications, Inc.

Carnap, R. (1961). On the use of Hilbert's ε-operator in scientific theories. In Bar-Hillel *et al.* Eds. (1961), 156–164.

Carnap, R. (1963a). Intellectual autobiography. In Schilpp Ed. (1963), pp. 3–84.

Carnap, R. (1963b). Replies and systematic expositions. In Schilpp Ed. (1963), pp. 859–1013.

Carnap, R. (1966). *Philosophical Foundations of Physics.* M. Gardner, Ed. New York: Basic Books.

Castonguay, C. (1971). Meaning and Existence in Mathematics. Ph. D. thesis, McGill University.

Castonguay, C. (1972). *Meaning and Existence in Mathematics.* Wien-New York: Springer-Verlag.

Chomsky, N. (1963). Formal properties of grammars. In Luce *et al.* 1963, Vol. II, pp. 323–418.

Chomsky, N. and G. A. Miller (1963). Introduction to the formal analysis of natural languages. In Luce *et al.* 1963, Vol. II, pp. 269–321.

Church, A. (1951). A formulation of the logic of sense and denotation. In P. Henle, H. M. Kallen and S. K. Langer, Eds., *Structure, Method and Meaning: Essays in Honor of Henry M. Sheffer*, pp. 3–24. New York: The liberal Arts Press.

Church, A. (1956). Propositions and sentences. In *The Problem of Universals: A Symposium*, pp. 3–11. Notre Dame, Ind.: University of Notre Dame Press.

Church, A. (1973/74). Outline of a revised formulation of the logic of sense and denotation. *Nous* **7**: 24–33, **8**: 135–156.

Chwistek, L. (1949). *The Limits of Science.* London: Routledge and Kegan Paul.

Cohen, L. J. (1962). *The Diversity of Meaning.* London: Methuen.

Cole, M. and I. Maltzman, Eds. (1969). *A Handbook of Contemporary Soviet Psychology.* New York and London: Basic Books, Inc.

Dummett, M. (1973). *Frege. Philosophy of Language.* London: Duckworth & Co. Ltd.

Feigl, H. (1950). Existential hypotheses. *Phil. Sci.* **17**: 35–62.

Feigl, H. (1958). The 'mental' and the 'physical'. In H. Feigl, M. Scriven, and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science*, II: 370–497.

Feyerabend, P. K. (1970). Against method: Outline of an anarchistic theory of knowledge. In Radner and Winokur Eds. (1970) pp. 17–130.

Fraassen, B. C. van (1971). *Formal Semantics and Logic.* New York: Macmillan.

Frege, G. (1879). Begriffsschrift. Repr. in *Begriffsschrift und andere Aufsätze.* I. Angelelli, Ed. Hildesheim: Georg Olms 1964.

Frege, G. (1891). Funktion und Begriff. In Angelelli (1967).

Frege, G. (1892). Über Sinn und Bedeutung. In Angelelli (1967).

Frege, G. (1894). Rezension von: E. G. Husserl, Philosophie der Arithmetik. I. In Angelelli (1967).

Frege, G. (1895). Le nombre entier. In Angelelli (1967).

Frege, G. (1912). Anmerkungen zu P. E. B. Jourdain, The Development of the Theories of Mathematical Logic, etc. In Angelelli (1967).

Frege, G. (1969). *Nachgelassene Schriften.* H. Hermes, F. Kambartel and F. Kaulbach, Eds. Hamburg: F. Meiner.

Freudenthal, H. (1971). More about *Foundations of Physics. Foundations of Physics* **1**: 315–323.

Geach, P. T. (1962). *Reference and Generality.* Ithaca, N.Y.: Cornell University Press.

Geach, P. T. and M. Black, Eds. (1952). *Translations from the Philosophical Writings of Gottlob Frege.* Oxford: Basil Blackwell.

Ginsburg, S. (1966). *The Mathematical Theory of Context-Free Languages.* New York: McGraw-Hill.

Goodman, N. (1951). *The Structure of Appearance.* Cambridge, Mass.: Harvard University Press.

Goodman, N. (1961). About. *Mind N.S.* **70**: 1–24.

Goodman, N. (1968). *The Languages of Art.* Indiana and New York: Bobbs-Merrill Co.

Harris, Z. (1968). *Mathematical Structure of Languages.* New York: Interscience.

Hartnett, W. E. (1963). *Principles of Modern Mathematics,* 1. Chicago: Scott, Foresman, and Co.

Hartnett, W. E. (1970). *Principles of Modern Mathematics,* 2. Chicago: Scott, Foresman, and Co.

Heisenberg, W. (1955). The development of the interpretation of the quantum theory. In W. Pauli, L. Rosenfeld, and V. Weisskopf, Eds., *Niels Bohr and the Development of Physics,* pp. 12–29. London: Pergamon.

Heisenberg, W. (1958). The representation of nature in contemporary physics. *Daedalus* **87**: 95–108.

Hempel, C. G. (1945). Studies in the logic of confirmation. *Mind N.S.* **54**: 1–26, 97–121.

Hempel, C. G. (1950). A note on semantic realism. *Phil. Sci.* **17**: 169–173.

Hempel, C. G. (1958). The theoretician's dilemma. In H. Feigl, M. Scriven, and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* II: 37–98. Minneapolis, Min.: University of Minnesota Press.

Hempel, C. G. (1965). *Aspects of Scientific Explanation and Other Essays in the Philosophy of Science.* New York: The Free Press.

Henkin, L. (1953). Some notes on nominalism. *J. Symbol. Logic* **18**: 19–29.

Henkin, L., P. Suppes and A. Tarski, Eds. (1959). *The Axiomatic Method.* Amsterdam: North-Holland Publ. Co.

Hilbert, D. (1900?). Letter to G. Frege. In Angelelli (1967).

Hilbert, D. and W. Ackermann (1950). *Principles of Mathematical Logic.* New York: Chelsea Publ. Co.

Hilbert, D. and P. Bernays (1968). *Grundlagen der Mathematik* I, 2nd ed. Berlin-Heidelberg-New York: Springer-Verlag.

Hill, T. E. (1971). *The Concept of Meaning.* New York: Humanities Press.

Hintikka, J. (1968). Varieties of information. In B. v. Rootselaar and J. F. Staal, Eds., pp. 311–332.

Hintikka, J. (1970). On semantic information. In W. Yourgrau and A. D. Breck, Eds., *Physics, Logic, and History,* pp. 147–172. New York and London: Plenum Press.

Katz, J. J. and J. A. Fodor (1963). The structure of a semantic theory. *Language* **39**: 170–210.

Kneale, W. and M. (1962). *The Development of Logic.* Oxford: Clarendon Press.

Kneale, W. (1972). Numbers and numerals. *Brit. J. Phil. Sci.* **23**: 191–206.

Kraft, V. (1970). *Mathematik, Logik and Erfahrung,* 2nd ed. Wien and New York: Springer-Verlag.

Kuhn, T. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Lambert, K. and B. C. v. Fraassen (1970). Meaning relations, possible objects, and possible worlds. In K. Lambert, Ed., *Philosophical Problems in Logic,* pp. 1–19. Dordrecht: D. Reidel Publ. Co.

Lawvere, F. W. (1966). The category of categories as a foundation of mathematics. *Proceedings of the Conference on Categorical Algebra* (La Jolla 1965), pp. 1–20. Berlin-Heidelberg-New York: Springer-Verlag.

Leonard, H. S. (1967). *Principles of Reasoning*. New York: Dover Publications Inc.

Lewis, C. I. (1944). The modes of meaning. *Philosophy and Phenomenological Research* **4**: 236–249.

Lewis, C. I. (1951). Notes on the logic of intensions. In P. Henle *et al.*, Eds., *Structure, Method and Meaning*. New York: Liberal Arts Press.

Linsky, L. (1967). *Referring*. London: Routledge and Kegan Paul.

Linsky, L., Ed. (1952). *Semantics and the Philosophy of Language*. Urbana: University of Illinois Press.

Luria, A. R. (1969). Speech development and the formation of mental processes. In Cole and Maltzmann, Eds. 1969, pp. 121–162.

Luce, R. D., R. R. Bush, and E. Galanter, Eds. (1963). *Handbook of Mathematical Psychology*, Vol. II. New York: John Wiley and Sons, Inc.

MacKay, D. M. (1969). *Information, Mechanism and Meaning*. Cambridge, Mass.: M.I.T. Press.

Marcus, S. (1967). *Algebraic Linguistics: Analytical Models*. New York: Academic Press.

Marhenke, P. (1950). The criterion of significance. In Linsky, Ed. (1952), pp. 139–159.

Montague, R. (1970). Pragmatics and intensional logic. *Synthese* **22**: 68–94.

Morris, C. (1938). *Foundations of the Theory of Signs*. Chicago: University of Chicago Press.

Nagel, E. (1950). Science and semantic realism. *Phil. Sci.* **17**: 174–181.

Nagel, E., P. Suppes, and A. Tarski, Eds. (1962). *Logic, Methodology and Philosophy of Science*. Stanford: Stanford University Press.

Neumann, J. v. (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: J. Springer.

Osgood, C. E., G. J. Suci, and P. H. Tannenbaum (1957). *The Measurement of Meaning*. Urbana, Ill.: University of Illinois Press.

Patton, T. E. (1965). Some comments on "About". *J. Phil.* **62**: 311–325.

Popper, K. R. (1963a). The demarcation between science and metaphysics. In Schilpp, Ed. (1963), pp. 183–226.

Popper, K. R. (1963b). *Conjectures and Refutations*. New York: Basic Books.

Popper, K. R. (1966). A theorem on truth-content. In P. K. Feyerabend and G. Maxwell, Eds., *Mind, Matter, and Method: Essays in Philosophy and Science in Honor of Herbert Feigl*, pp. 343–353. Minneapolis: University of Minnesota Press.

Przelecki, M. (1969). *The Logic of Empirical Theories*. London: Routledge and Kegan Paul.

Putnam, H. (1970). Is semantics possible? *Metaphilosophy* **1**: 187–201.

Putnam, H. (1971). *Philosophy of Logic*. New York: Harper & Row Publ.

Putnam, H. (1973). Meaning and reference. *Journal of Philosophy* **70**: 699–711.

Quine, W. V. (1952). *From a Logical Point of View*. Cambridge, Mass.: Harvard University Press.

Quine, W. V. (1966). *The Ways of Paradox and Other Essays*. New York: Random House.

Quine, W. V. (1971). Epistemology naturalized. *Proc. XIVth Intern. Congress Phil.* VI: 87–103. Wien: Herder.

Radner, M. and S. Winokur, Eds. (1970). *Minnesota Studies in the Philosophy of Science* IV. Minneapolis: University of Minnesota Press.

Robinson, A. (1963). *Introduction to Model Theory and to the Metamathematics of Algebra*. North-Holland Publ. Co.

Robinson, A. (1966). *Non-standard Analysis*. North-Holland Publ. Co.

Rootselaar, B. van, and J. F. Staal, Eds. (1968). *Logic, Methodology and Philosophy of Science* III. Amsterdam: North-Holland Publ. Co.

Rosen, R. (1958). The representation of biological systems from the standpoint of the theory of categories. *Bull. Math. Biophys.* **20**: 317–341.

Rozeboom, W. (1962). The factual content of theoretical concepts. In H. Feigl and G.

Rosenbloom, P. (1950). *The Elements of Mathematical Logic.* New York: Dover.

Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* III: 273–357. Minneapolis, Min.: University of Minnesota Press.

Rozeboom, W. (1970). The crisis in philosophical semantics. In Radner and Winokur, Eds. pp. 196–219.

Russell, B. (1918). The philosophy of logical atomism. *The Monist* **28**: 495–527; **29**: 32–63, 190–222, 345–380.

Russell, B. (1940). *An Inquiry Into Meaning and Truth.* London: George Allen & Unwin.

Russell, B. (1948). *Human Knowledge: Its Scope and Limits.* London: George Allen & Unwin.

Russell, B. (1959). *My Philosophical Development.* London: George Allen & Unwin.

Ryle, G. (1949). *The Concept of Mind.* London: Hutchinson.

Schaff, A. (1962). *Introduction to Semantics.* Oxford: Pergamon Press.

Schilpp, P. A., Ed. (1963). The Philosophy of Rudolf Carnap. La Salle, Ill.: Open Court; London: Cambridge University Press.

Scott, D. (1970). Advice on modal logic. In K. Lambert, Ed., *Philosophical Problems in Logic.* Dordrecht: D. Reidel Publ. Co.

Searle, J. R. (1969). *Speech Acts: An Essay in the Philosophy of Language.* Cambridge: University Press.

Seshu, S. and M. B. Reed (1961). *Linear Graphs and Electrical Networks.* Reading, Mass.: Addison-Wesley.

Shannon, C. and W. Weaver (1949). *The Mathematical Theory of Communications.* Urbana, Ill.: University of Illinois Press.

Sharvy, R. (1972). Three types of referential opacity. *Phil. Sci.* **39**: 152–161.

Shwayder, D. S. (1961). *Modes of Referring and the Problem of Universals.* Berkeley and Los Angeles: University of California Press.

Suppe, F. (1971). On partial interpretation. *J. Phil.* **68**: 57–76.

Suppes, P. (1960). *Axiomatic Set Theory.* Princeton, N.J.: D. Van Nostrand.

Suppes, P. (1967). *Set-theoretical Structures in Science.* Stanford University: Institute for Mathematical Studies in the Social Sciences.

Suppes, P. (1969). *Studies in the Methodology and Foundations of Science.* Dordrecht: D. Reidel.

Suszko, R. (1967). An essay in the formal theory of extension and intension. *Studia Logica* **20**: 7–28.

Tarski, A. (1956). *Logic, Semantics, Metamathematics.* Oxford: Clarendon Press.

Tarski, A., A. Mostowski, and R. M. Robinson (1953). *Undecidable Theories.* Amsterdam: North-Holland Publ. Co.

Vaihinger, H. (1920). *Die Philosophie des Als Ob*, 4th ed. Leipzig: F. Meiner.

Waddington, C. H. (1967). Discussion of Eden's paper. In P. S. Moorhead and M. M. Kaplan, Eds., *Mathematical Challenges to the Neo-Darwinian Interpretation of Evolution.* Philadelphia: The Wistar Institute.

Watson, W. H. (1950). *On Understanding Physics.* New York: Harper & Brothers.

Weingartner, P. (1972). Die Fraglichkeit der Extensionalitätsthese und die Probleme einer intensionalen Logik. In R. Haller, Ed., *Jenseits von Sein und Nichtsein.* Graz: Akademische Druck-u. Verlagsanstalt.

Weingartner, P. (1973). A predicate calculus for intensional logic. *J. Phil. Logic* **2**: 58–141.

Whitehead, A. N. and B. Russell (1927). *Principia Mathematica* I. 2nd. ed. Cambridge: Cambridge University Press.

Wigner, E. P. (1970). Physics and the explanation of life. *Foundations of Phys.* **1**: 35–45.

Williams, D. C. (1937). The realistic interpretation of scientific sentences. *Erkenntnis* **7**: 169–178, 375–382.

Williams, D. C. (1966). *Principles of Empirical Realism.* Springfield, Ill.: Charles C. Thomas.

Williams, G. C. (1966). *Adaptation and Natural Selection.* Princeton, N.J.: Princeton University Press.

Williams, M. B. (1970). Deducing the consequences of evolution: A mathematical model. *J. Theoret. Biol.* **29**: 343–385.

Wojcicki, R. (1966). Semantical criteria of empirical meaningfulness. *Studia Logica* **19**: 75–102.

Yourgrau, W. and A. D. Breck, Eds. (1970). *Physics, Logic, and History.* New York and London: Plenum Press.

Zinov'ev, A. A. (1973). *Foundations of the Logical Theory of Scientific Knowledge (Complex Logic).* Dordrecht: D. Reidel.

INDEX OF NAMES

Ackermann, W. 66
Agassi, Joseph 114
Ajdukiewicz, Kazimierz 14
Anderson, Alan Ross 65
Angel, Roger B. xiii, 35
Arbib, Michael A. 11
Aristotle 33, 65
Arnauld, Antoine 117

Beatty, Harry xiii
Belarmino, Cardinal 97
Bell, J. L. 33
Bar-Hillel, Yehoshua 136
Bernays, Paul 63
Blumenthal, L. M. 131
Bolzano, Bernard 17, 28, 32, 33, 117
Bourbaki, Nicholas 22, 119
Boutroux, Emile 89
Braithwaite, Richard B. 161–162
Brillouin, Léon 69
Bunge, Mario 5, 35, 38, 40, 58, 67, 70, 75, 80, 111, 113, 149, 151, 161, 171
Buridan, Jean 49

Carnap, Rudolf 4, 48, 117, 122, 134, 136, 138, 161–162, 164
Castonguay, Charles xiii, 117, 124, 150
Chomsky, Noam 1, 11
Church, Alonzo 17, 22
Chwistek, Leon 20
Cohen, L. Jonathan 3
Corcoran, John xiii

Eddington, Arthur Stanley 20

Feigl, Herbert 43, 160, 162
Felscher, Walter xiii
Feyerabend, Paul K. 67–68, 78
Frege, Gottlob 4, 16–17, 20, 22, 24, 47–48, 123
Freudenthal, Hans 35

Galilei, Galileo 97
Geach, Peter 49
Ginsburg, S. 11
Goodman, Nelson 30, 81

Hamilton, William 117
Harris, Zelig 11
Hartnett, William E. 4
Heisenberg, Werner 34
Hempel, Carl G. 79, 138, 160–162
Henkin, Leon 13
Hilbert, David 20, 63, 66, 170
Hill, Thomas English 3
Hintikka, Jaako 136
Hume, David 81, 89
Husserl, Edmund 13, 48

James, William 81

Kant, Immanuel 89
Kessler, Glenn xiii
Kirschenmann, Peter xiii
Kleiner, Scott A. xiii
Kneale, Martha 117
Kneale, William 13, 117
Kolmogoroff, Aleksander N. 137
Kraft, Viktor 85
Kuhn, Thomas 67, 78
Kurosaki, Hiroshi xiii

Lambek, Joachim xiii
Lambert, Karel 140
Lawvere, F. William 119
Leibniz, Gottfried Wilhelm 24
Leonard, Henry 124
Lewis, Irving Clarence 117, 122
Linsky, Leonard 22, 24
Lungarzo, Carlos Alberto xiii
Luria, A. R. 3

MacKay, David M. 104
Marcus, Solomon 11
Marhenke, Paul 7
Miller, George Armitage 1, 11
Meinong, Alexius 117
Menger, Karl 131
Montague, Richard 171
Morris, Charles 29

Nagel, Ernest 160
Negrepontis, Stelios xiii
Neumann, John von 70
Nicole, Pierre 117
Nuño, Juan A. xiii

Ockham, William 20
Oppacher, Franz xiii
Osgood, Charles 3

Patton, Thomas 81
Peirce, Charles Sanders 89, 138
Popper, Karl R. 117, 134, 136, 138, 139, 159
Przelecki, Marian 161
Probst, David xiii
Putnam, Hilary 20

Quine, Willard Van Orman 20, 123, 138, 171

Reed, M. B. 102
Robinson, Abraham 3, 20
Rosen, Robert 102
Rosenbloom, Paul C. 34
Rozeboom, William W. 160–161
Russell, Bertrand 17, 85, 120, 138
Ryle, Gilbert 79

Salt, David xiii
Schaff, Adam 3
Seshu, S. 102
Scott, Dana 171
Searle, John R. 7
Shakespeare, William 22
Shannon, Claude 137
Sharvy, Richard 123
Shwayder, D. S. 53
Simon, Simple 136
Slomson, A. B. 3
Soran, Somnez xiii
Suppe, Frederick 162
Suppes, Patrick 120
Suszko, Roman 124

Tammelo, Ilmar xiii
Tarski, Alfred 4, 20, 39, 63
Törnebom, Håkan 113
Torretti, Roberto xiii
Tuomela, Raimo xiii

Vaihinger, Hans 13
Van Fraassen, Bas C. 3, 140
Vienna Circle 138

Wang, Hao 146
Watson, W. H. 104
Weaver, Warren 137
Weingartner, Paul xiii, 124
Whitehead, Alfred North 120
Williams, Donald 138
Williams, George 33
Williams, Mary B. 149
Wittgenstein, Ludwig 48, 133, 156
Wojcicki, Ryszard 161

Zinov'ev, A. A. 29

INDEX OF SUBJECTS

Analogy 103
Analytic 55–56. *See also* Tautology
Assumption, semantic 104–114
Axiom 149

Background of a theory 151–152
Boolean algebra 144, 151

Category, Aristotelian 131
  mistake 79
  semantic 13–14
Closure, semantical 65
Coextensiveness 134–135
Cointensiveness 125
Commensurability of theories, methodological 77–78
  referential 66–68, 77–78
Communication 8
Complexity, semantic 169
Concept 14–18
  factual 59
  formal 60
  modal 122–123
  pragmatic 121–122
Confirmation 140
  paradox of 79–80
Construct 13, 31, 43, 57, 88, 92
Content 158–166
  changes in 166
  empirical 161–163
Context 57, 143
  closed 143–145
  factual 60
  formal 60
Conventionalism 70–72, 160–161
Core sense 169
Coreference 57–58
Correspondence rule 25, 106. *See also* Denotation rule, Semantic assumption
Craig's interpolation theorem 66
  reduction 164

Data 70
Denotation 25, 42–43
  rule 25, 105–109
Designation 11, 21–25, 42
  function 23–25
  rule 25
Difference in gist 154
  intensional 130–133
Differentia specifica 130

Empiricism 113. *See also* Operationism, Phenomenalism, Positivism
Entailment 128–129, 145
Equivalence, logical 53, 130
  of representations 95
Evidence 44–45, 163
  relevant 79
Expression 10–11, 14, 24, 26
Extension 34–35, 49, 118–120, 130. *See also* Volume 2, Chapter 9, Sec. 1
Extensionalism 118–120

Family resemblance 133
Filter 155–156
Form, logical 115–116
Formalism, mathematical 104–105
Foundations of a theory 152

Gist 147, 149–154
  changes in 153–154
  extralogical 151–153
Grammar 11, 46
Grammaticalism 46

Holism, semantic 152, 171
Hypothesis 111

Ideal 145–146
Import 116–117, 134, 142, 156–165
  extralogical: see Content
Information, semantic 121, 135–138

  statistical 137–138
Intension 115–141, 170
Intensional context 121–122
Interpretation, factual 105–114. *See also* Volume 2, Chapter 6
  mathematical 105

Language 8–13, 18–20
  conceptual 23–26
Lattice 143–145, 155
Law statement 89, 94
Logic 4–5, 39, 64
  intensional 121

Magnitude 95–97
Materialism, conceptualist 32
Mathematics 4–5, 39, 120
Meaning 3, 6, 29, 45, 109, 172. *See also* Intension, Reference, Sense, and Volume 2, Chapter 7
  change 153–154
  surplus 160
Metaphysics 26–30, 40, 42, 52–53, 81. *See also* Ontology
Metastatement 35
Metatheory 4
Methodology 45, 139–140
Modalities 141
Model, theoretical 99–100. *See also* Theory, specific
Model object: *see* Schema
Model theory 3, 39
Modeling 99–104
Monoid 10

Name 21–23
Negation 85–86
Nominalism 20–22, 29–32

Object 26–27
Observer 72–73
Ontology 26, 38, 39. *See also* Metaphysics
Operationism 47, 109, 160–165. *See also* Positivism

Philosophy, exact 5
Phenomenalism 71–75
Platonism 28–31
Positivism 160
Pragmatics 12, 71, 98
Pragmatism 47, 71
Predicate 15–17, 50
  maximal 60–61
Presupposition 151
Primitive 149
Probability 69, 86, 90–91, 109–110, 136
Proposition 8, 13, 15, 17–18, 28. *See also* Statement
Propositional function 15. *See also* Predicate
Proxying relation 91–93
Pseudopredicate 16
Psycholinguistics 1–2
Psychologism 46
Psychology 69–70
Purport 116–117, 142, 146–148
  extralogical 147–148

Quantum mechanics 73–74, 111–113
Quotation marks 23

Ramsey sentence 164
Realism 71–75, 109
  empirical 160–161
Reference 32–82, 119
  class 37–80
  deep 52
  factual 39–42, 59–75
  functions 49–52
  genuine 68–70
  immediate 36
  mediate 36
  oblique 52
  ostensive 52
  relation 25, 34–43, 86–87
  spurious 68–70
Referential heterogeneity 58
  homogeneity 58
  partition 58
Relativity theory 58, 73–74
Relevance 75–77
  principle of 66
Representation, conceptual 82–114
  accurate 90–91
  assumption 105–114
  equivalent 94–95, 97–98, 101–102
  relation 87–93

Ring of intensions 132–133

Scale concept 40–41
Schema 99–101
Semantics ix, 1–7, 11–12, 45–46, 68, 96
Sense 5, 116–117, 142–172
Sentence 10, 14, 17–20, 27–28
Set theory 120
Sign 8, 27, 31
Statement 16, 19, 51–52. *See also* Proposition
Strength, logical 134
String 10, 30. *See also* Term, Expression
Subjectivism 109
Syntax 10
System, semantic 25

Tautology, intension of 128
  reference of 54–56
  sense of 168
Term 14, 26
Testability 138–140
Theoretical entity 71
Theory 18–20, 83
  axiomatic 62–67
  content 158–165
  exposure 77–78
  factual 61–75, 105
  generic 99–100
  reference 62–75
  rival 78–79
  semantic components 104–114
  semantically equivalent 98
  specific 99–100
  universe of discourse 62
  versatility 160
  vibrating string 92
Topology in intension space 133
Transformation formula 97–98
Truth 39, 47–48, 159. *See also* Volume 2, Chapter 8
  condition 12

Wff. *See* Expression

The companion of this book is Volume 2 of the

Treatise on Basic Philosophy

INTERPRETATION AND TRUTH

TABLE OF CONTENTS

PREFACE

6 Interpretation

7 Meaning

8 Truth

9 Offshoots

10 Neighbors

INDEX OF NAMES

INDEX OF SUBJECTS

BIBLIOGRAPHY

TREATISE ON BASIC PHILOSOPHY

Volume 2

SEMANTICS II: INTERPRETATION AND TRUTH

TREATISE ON BASIC PHILOSOPHY

1

SEMANTICS I *Sense and Reference*

2

SEMANTICS II *Interpretation and Truth*

3

ONTOLOGY I *The Furniture of the World*

4

ONTOLOGY II *A World of Systems*

5

EPISTEMOLOGY I *The Strategy of Knowing*

6

EPISTEMOLOGY II *Philosophy of Science*

7

ETHICS *The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 2

Semantics II:

INTERPRETATION AND TRUTH

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND / BOSTON-U.S.A.

Library of Congress Catalog Card Number 74–83872

ISBN-13 978-90-277-0573-0 e-ISBN-13 978-94-010-9922-6
DOI: 10.1007/978-94-010-9922-6

Published by D. Reidel Publishing Company,
P.O. Box 17, Dordrecht, Holland

Sold and distributed in the U.S.A., Canada, and Mexico
by D. Reidel Publishing Company, Inc.
306 Dartmouth Street, Boston,
Mass. 02116, U.S.A.

All Rights Reserved
Copyright © 1974 by D. Reidel Publishing Company, Dordrecht, Holland
No part of this book may be reproduced in any form, by print, photoprint, microfilm,
or any other means, without written permission from the publisher

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and of right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient – except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: "Do your own thing. Your reward will be doing it, your punishment having done it".

CONTENTS OF *SEMANTICS II*

PREFACE XI

SPECIAL SYMBOLS XIII

6. INTERPRETATION 1
   1. Kinds of Interpretation 1
   2. Mathematical Interpretation 3
      - 2.1. Abstract Theory 3
      - 2.2. Model 5
      - 2.3. Intensional Models and Extensional Models 8
      - 2.4. Insufficiency of Extensional Models 9
   3. Factual Interpretation 13
      - 3.1. The Need for Factual Interpretation in Science 13
      - 3.2. How Interpretations are Assigned and What They Accomplish 15
      - 3.3. The Factual Interpretation Maps 18
      - 3.4. Factual Interpretation: Full and Partial 21
      - 3.5. Generic Partially Interpreted Theories 24
      - 3.6. Principles of Factual Interpretation 26
      - 3.7. Factual Interpretation and Truth 29
      - 3.8. Interpretation and Exactification 32
   4. Pragmatic Aspects 35
      - 4.1. Pragmatic Interpretation 35
      - 4.2. The Interpretation Process 37
   5. Concluding Remarks 40

7. MEANING 42
   1. Babel 42
   2. The Synthetic View 45
      - 2.1. Meaning as Sense cum Reference 45
      - 2.2. Significance 50
      - 2.3. Significance Assignment 53
      - 2.4. Degrees of Significance Definiteness 55
   3. Meaning Invariance and Change 57
      - 3.1. Synonymy 57
      - 3.2. Meaning Invariance 59
      - 3.3. Meaning Change 62

4. Factual and Empirical Meanings 64
4.1. Definitions 64
4.2. The Search for Factual Meaning 65
4.3. Shape and Role of Meaning Assumptions 68
5. Meaning *et alia* 72
5.1. Meaning and Testability 72
5.2. Meaning and Use 74
5.3. Meaning and Understanding 75
5.4. Factual Meaning and Covariance 76
6. Concluding Remarks 80

8. TRUTH 81

1. Kinds of Truth 82
1.1. Truth Bearers 82
1.2. Truth Values: Acquired 86
1.3. Quadruple Truth 87
2. Truth of Reason and Truth of Fact 90
2.1. Truth of Reason 90
2.2. Truth of Fact: The Synthetic View 93
2.3. Truth Values: Conditional 97
2.4. Truth Conditions 101
3. Degrees of Truth 105
3.1. The Problem and How to Fail to Solve It 105
3.2. Axioms 108
3.3. Topologies of S_D 110
3.4. Comparing Truth Values 113
3.5. Scientific Inference 117
3.6. Comments 118
4. Truth *et alia* 123
4.1. Truth and Probability 123
4.2. Truth, Meaning, and Confirmation 125
4.3. Truth and Belief 127
4.4. Truth and Time 129
5. Closing Remarks 130

9. OFFSHOOTS 133

1. Extension 133
1.1. The Problem 133
1.2. Strict Extension: Definition 135
1.3. Some Consequences 137
1.4. Comparing Extensions 140
1.5. Algebraic Matters 142
1.6. Extension and Intension: the Inverse Law 144
1.7. Concluding Remarks 146

2. Vagueness 147
    2.1. Meaning Vagueness 147
    2.2. Extensional Vagueness 150
    2.3. Structural Indefiniteness 152
3. Definite Description 153
    3.1. The Received View: Criticism 153
    3.2. An Elementary Analysis of Definite Descriptions 155
    3.3. A Mathematical Analysis of Definite Descriptions 158
    3.4. Continuation of the Analysis 160
    3.5. Meaning Questions 161
    3.6. Truth Questions 163
    3.7. The Real Size of the Theory of Descriptions 164

10. NEIGHBORS 166

1. Mathematics 166
    1.1. The Relevance of Semantics to Mathematics 166
    1.2. On Extensionalism 167
    1.3. On Objectivity 169
2. Logic 170
    2.1. Analyticity 170
    2.2. Definition 174
    2.3. Presupposition 177
3. Epistemology 179
    3.1. The Status of Epistemology 179
    3.2. Representation vs. Instrument and Picture 180
    3.3. Objectivity vs. Subjectivity 183
    3.4. The Knowing Subject 186
4. Metaphysics 189
    4.1. The Metaphysical Neutrality of Language 189
    4.2. The Metaphysical Neutrality of Logic 190
    4.3. Metaphysical Commitments of the Semantics of Science 194
5. Parting Words 195

BIBLIOGRAPHY 198

INDEX OF NAMES 206

INDEX OF SUBJECTS 208

PREFACE TO *SEMANTICS II*

This is the second and last part of our work on semantics. The first part, titled *Sense and Reference*, constitutes volume 1 of the *Treatise*.

What follows presupposes an understanding of the slippery notions of sense and reference. Any theories elucidating these concepts will do for the purpose of tackling the present volume. But, of course, only the theories expounded in Part I will articulate cogently with those we are about to discuss. Nevertheless, the gist of Part I can be summarized in a few words.

Philosophical semantics is about constructs, in particular predicates and propositions. Every such object has both a sense and a reference. The full sense of a construct is the collection of its logical relatives. This collection is made up of two parts: the purport, or set of implicants, and the import, or set of implicates. For example, the purport of a defined concept is the set of concepts entering in its definition, and its import is the collection of concepts hanging from it. As to the referents of a predicate, they are the individuals occurring in its domain of definition. And the reference class of a statement is the union of the reference classes of all the predicates occurring in the proposition. Some constructs, notably those occurring in ordinary knowledge and in scientific theories, have a factual sense and a factual reference. The theories of sense and reference put forth in Part I allow one to calculate both the sense (in particular the factual sense) and the reference class (in particular the factual reference class) of any predicate and any statement. By this token they can help us solve certain tricky semantical problems posed by some of the most important scientific theories, the sense and reference of which are often the object of spirited debates. So much for a résumé of Part I.

The present volume starts with the problem of interpretation. Interpretation is construed as the assignment of constructs (e.g. predicates) to symbols. It can be purely mathematical, as when the dummy x is interpreted as an arbitrary natural number, or also factual, as when such a number is interpreted as the population of a town. Now, as we saw a while ago, predicates and propositions have both a sense and a reference – and nothing else as far as meaning is concerned. These, then, are taken to be the meaning components. That is, the meaning of a construct is defined as the ordered couple constituted by its sense and its reference class. Once the meaning of a proposition has been established we can proceed to finding out its truth value – provided it has one. If the proposition happens to be factual, i.e. to have factual referents, then it may be only partially true – if true at all. Hence we must clarify the concept of partial truth of fact. This we do by building a theory that combines features of both the correspondence and the coherence theories of truth. The remaining semantical notions, notably those of extension, vagueness, and definite description, are made to depend on the concepts of meaning and truth and are therefore treated towards the end of this work. The last chapter explores the relations between philosophical semantics and other branches of scholarship, in particular logic and metaphysics.

This volume, like its predecessor, has been conceived with a definite goal, namely that of producing a system of philosophical semantics capable of shedding some light on our knowledge of fact, whether ordinary of scientific. We leave the semantics of natural languages to linguists, psycholinguists and sociolinguists, and the semantics of logic and mathematics (i.e. model theory) to logicians and mathematicians. Our central concern has been, in other words, to clarify and systematize the notions of meaning and truth as they occur in relation to factual knowledge. For this reason our semantics borders on our epistemology.

SPECIAL SYMBOLS

C	Set of constructs (concepts, propositions, or theories)
$\mathbb{C}$	Context
$\mathscr{C}$	Content (extralogical import)
$\mathscr{C}n$	Consequence
$\mathscr{D}$	Designation
Δ	Denotation
$\hat{=}$	Representation
$\mathscr{E}$	Extension
$\mathscr{I}$	Intension
$\mathscr{I}mp$	Import (downward sense)
L	Logic
$\mathscr{L}$	Language
$\mathscr{M}$	Meaning
Ω	Universe of objects (of any kind)
$\mathbb{P}$	Family of predicates
$\mathscr{P}ur$	Purport (upward sense)
$\mathscr{R}$	Reference
S	Set of Statements (propositions)
$\mathscr{S}$	Sense
$\mathscr{S}ig$	Signification
T	Theory (hypothetico-deductive system)
$\mathscr{V}$	Truth value function

CHAPTER 6

INTERPRETATION

All of the symbols in a scientific theory are interpreted. They are interpreted as designating certain mathematical concepts, some of which are in turn interpreted as representing certain aspects of the world. Such a double interpretation must be shown as completely and explicitly as possible if the signification of the symbolism is to emerge clearly. But what is an interpretation, in particular a factual one? This is the central theme of the present chapter.

1. KINDS OF INTERPRETATION

Anything, from sign to gait, can be interpreted if one knows how to. Thus farmers interpret cloud shapes, physicians bodily appearances, and charlatans dreams. In all three cases observed facts are correlated to hypothesized ones and the latter are assumed to explain the former. This kind of interpretation, bearing on natural signs, may be called *epistemic*: actually it is a mode of explanation. The concern of semantics is with another kind of interpretation, one bearing either on signs or on constructs. Henceforth we shall adopt this acceptation of 'interpretation', which may be called *semiotic.*

Semiotic interpretation may be construed either as bearing on signs or as bearing on constructs. The interpretation of signs is a task for designation rules, while the interpretation of constructs is performed by semantic assumptions. Example of sign interpretation: '&' *designates* conjunction. Example of construct interpretation: $F(a, b)$ *represents* the strength of the interaction between a and b.

Whether bearing on signs or on constructs, interpretation is required whenever that which is interpreted is not definite enough. Interpretation goes from the less to the more definite or specific. For example, from an ambiguous sign like 'S' to a definite generic construct such as "set", from the latter to a specific construct like "the set of pairs", or from here to a factual item like the collection of married couples, or to an empirical item such as the collection of married couples counted by the census bureau. We distinguish then four different kinds of interpretation relations, that are exhibited and exemplified in Table 6.1.

TABLE 6.1

Semiotic interpretations

Kind of interpretation		Relata	Example
1 Designation	$\mathscr{D}$	Symbol→Construct	Function letter→Function
2 Mathematical	μ	Generic construct→Specific construct	Function→sin
3 Factual	ϕ	Specific construct→Factual item	$\sin \omega t$→pendulum elongation
4 Pragmatic	π	Specific construct→Empirical item	$\sin \omega t$→measured value of pendulum elongation

The first kind of interpretation, i.e., designation, occurs in every conceptual system: without designation rules a symbolism does not symbolize. Thus a page of the *Journal of Mathematical Psychology* may be regarded as a system of conventional signs (words and mathematical symbols) together with a set of interpretation conventions – mostly tacit but nevertheless operative. In other words, a conceptual system may be regarded as an *interpreted language*, i.e., a symbolism together with a collection of designation rules. An uninterpreted language, i.e., a well constructed system of artificial signs with no designata, would be as idle and unintelligible as a scientific manuscript after a total nuclear holocaust. The very notion of a totally uninterpreted language makes no sense except for purposes of analysis.

The most basic of all conceptual systems are of course the logical calculi: they are the most abstract in the sense that they are the least interpreted. Logical calculi are prime examples of *abstract theories*, i.e. of theories containing predicates with no fixed interpretation – hence making room for a variety of interpretations. But they are all interpreted languages in the sense that they contain a designation rule for every sign type. Thus a predicate letter like '*P*' is interpreted as an arbitrary predicate or attribute. The interpretation is limited to designation: the calculus is uninterpreted solely in the sense that it involves none of the kinds 2 to 4 listed in Table 6.1 above. Hence the calculus fails to characterize its individuals and to assign them definite properties: it deals with unspecified individuals and attributes. Ergo it contains no specific laws, i.e., laws satisfied by objects of a definite kind, such as earthquakes or revolutions. In short the predicate calculus, by being semantically noncommittal, is attached to no ontology. But it is not an empty symbolism either: its small case letters are interpreted as unspecified individuals, its capital letters as unspecified predicates, and so on.

The indicated interpretation of logical calculi in terms of constructs of a certain type is the usual or *standard* interpretation but not the only possible one. Logical calculi can be assigned alternative interpretations – but then they may cease to be logical theories, i.e., theories concerned with deductive inference. A well known *nonstandard* model of the propositional calculus is the one in terms of switches in an electric network. And Kolmogoroff's interpretation of the intuitionistic propositional calculus in terms of problems is a nonstandard model of that calculus. These examples are mentioned only as a reminder that logical calculi are abstract theories except for the designation rules (e.g., ⌜'p' designates a proposition⌝), which we do not always care to state explicitly.

2. MATHEMATICAL INTERPRETATION

2.1. *Abstract Theory*

The mathematical theories occurring in factual science, such as trigonometry and the infinitesimal calculus, come with a definite mathematical interpretation. In other words they are specific ("concrete") theories concerned with mathematical objects of a definite kind, such as plane triangles or real functions. Thus the formulas $\ulcorner \sin^2 x + \cos^2 x = 1 \urcorner$ and $\ulcorner d \sin x/\mathrm{d}x = \cos x \urcorner$ are uniquely interpreted, namely in the field of real numbers. The latter can be extended to the field of complex numbers, but this is yet another specific structure: it is just an example or model of a field.

By contrast to such fully interpreted theories, those in logic, abstract algebra and topology are calculi with no fixed sense over and above the one determined by their axioms. These calculi, or *abstract theories* as we prefer to call them, are sometimes called *languages* or even *uninterpreted languages*. But this name seems misleading. First because, unlike a language, a theory, no matter how abstract, contains definite assumptions (axioms). Second because these assumptions give the theory a definite if parsimonious sense: it may be called the *minimal sense* of any theory built upon the given abstract theory by interpreting or specifying some or all of its concepts. Such a further interpretation will turn the abstract theory into a "concrete" theory with a richer sense – and correspondingly a narrower extension. The consideration of an example will clarify these points.

Consider lattice theory L. This is an abstract or formal theory concerned with a roomy structure $\mathscr{L} = \langle S, \preccurlyeq, \wedge, \vee \rangle$ that fits a number of species of specific mathematical objects. Being thus unattached to any specific interpretation, lattice theory can espouse (and subsequently divorce) a number of alternative interpretations. These are interpretations of a mathematical theory within mathematics: they are *mathematical interpretations*. And these are superimposed on the designation rules that turn a symbolism into an abstract theory – in this case L. A few such additional (or mathematical) interpretations of L are listed in Table 6.2.

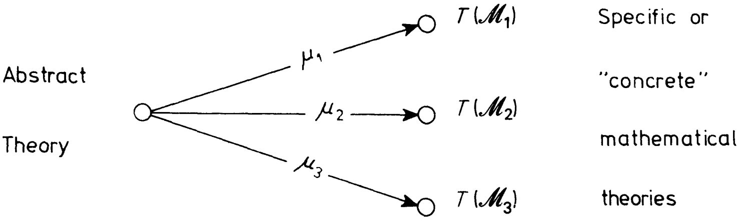
TABLE 6.2
Four mathematical interpretations of lattice theory

Primitives of L	Order interpretation	Class interpretation	Propositional interpretation	Arithmetic interpretation
Abstract set S	Abstract set S	A collection F of abstract sets	The set P of propositions	The set N of natural numbers
Partial order $\preccurlyeq$	Partial order $\preccurlyeq$	Set inclusion $\subseteq$	Entailment $\vdash$	Divisibility ⁒
Binary operation $\wedge$	Greatest lower bound	Set intersection $\cap$	Conjunction &	Greatest common divisor
Binary operation $\vee$	Least upper bound	Set union $\cup$	Disjunction $\vee$	Least common divisor

Table 6.2 illustrates the following important points.

(i) Mathematical interpretation is a construct-construct relation and, more particularly, an *intertheoretical* affair. This contrasts with the other three kinds of interpretation listed in Table 6.1.

(ii) Mathematical interpretation is a *one-many relation* between the set of abstract theories and the set of "concrete" (specific) theories:



(iii) Not all of the interpretations of an abstract theory are equally "concrete" or *specific*. For example, the order interpretation of L maps S and $\preccurlyeq$ onto themselves and specifies only $\wedge$ and $\vee$. Thus the referents remain nearly as indeterminate as before. The class interpretation of L is more "concrete" or familiar but not completely so: the domain F could in turn be interpreted by specifying the nature of the sets in F. Only the propositional and the arithmetic interpretations are full, i.e., not susceptible to any further specification – except of course instantiation, as when a definite proposition is picked out of the set P.

(iv) Every specific structure, such as $\mathscr{F} = \langle F, \subseteq, \cap, \cup \rangle$ or $\mathscr{P} = \langle P, \vdash, \&, \vee \rangle$, is a realization or *model* of the abstract structure $\mathscr{L} = \langle S, \preccurlyeq, \wedge, \vee \rangle$. That is, the new (specific) constructs satisfy the formulas of the abstract theory L. Equivalently: the formulas of the abstract theory are satisfied in any of its models.

(v) A specific structure, or model, may be regarded as the value of an interpretation function μ that maps abstract primitives into specific ones. Examples:

$$\mathscr{L} \overset{\mu_1}{\mapsto} \mathscr{F}, \qquad \mathscr{L} \overset{\mu_2}{\mapsto} \mathscr{P}.$$

Let us take a closer look at models – in the mathematical or model theoretic sense of the word not in any of the alternative senses. (For the various senses of 'model' see Bunge (1973a).)

2.2. *Model*

An axiomatized theory may be construed as describing the structure constituted by its specific basic or primitive concepts. Thus the general theory T of partial order may be regarded as the theory about the abstract relational structure $\mathscr{A}=\langle S, \preccurlyeq\rangle$, where S is an arbitrary set and $\preccurlyeq$ an ordering of S. Since neither S nor $\preccurlyeq$ is definable in T, they are primitives of T. And since they are mutually independent, as well as sufficient to develop T provided some logic has been presupposed, $\mathscr{A}$ is the primitive base of T. Equivalently: T is the theory of $\mathscr{A}$, or $T(\mathscr{A})$ for short.

Let us emphasize the abstract character of $\mathscr{A}$. The elements of S are utterly faceless, hence $\preccurlyeq$ itself is quite anonymous except for the axioms of T that determine the sense of $\preccurlyeq$ – i.e., the properties of reflexivity, antisymmetry, and transitivity. This, then, is the basic purport or gist of T – that S is a partially ordered set. (Recall Ch. 5, Sec. 3.3.) It would be nonsense to say that T is meaningless. The axioms of T provide the *minimal sense* of any theory obtained by assigning to S and $\preccurlyeq$ a specific interpretation i.e., by exemplifying both primitives of T.

Take now any of the specific mathematical theories resulting from assigning both S and $\preccurlyeq$ definite senses within mathematics. Consider, in particular, the propositional model $\mathscr{P}$ and the real number model $\mathscr{R}$ of the abstract structure $\mathscr{A}=\langle S, \preccurlyeq\rangle$:

(I_1) $\mu_1(S)=$The set P of propositions, $\mu_1(\preccurlyeq)=$The relation $\vdash$ of entailment,

(I_2) $\mu_2(S)=$The set R of real numbers, $\mu_2(\preccurlyeq)=$The smaller than or equal to relation $\leqslant$.

The outcome of each interpretation of the primitive of T is a specific relational structure or *model*:

$$\mathscr{M}_1=\mathscr{P}=\langle P, \vdash\rangle, \qquad \mathscr{M}_2=\mathscr{R}=\langle R, \leqslant\rangle.$$

These are models or realizations of the abstract structure $\mathscr{A}=\langle S, \preccurlyeq\rangle$. Since the axioms of the abstract theory $T(\mathscr{A})$ are satisfied under either interpretation, they are said to hold (or to be true) in the corresponding model. Upon adjoining either interpretation assumption (or semantic formula) to T, we get a "concrete" (specific) theory i.e. one concerned with a definite species of objects, such as propositions or real numbers. The object of such an interpreted theory being a model or specific structure, we may call the former the *theory of the model*, or $T(\mathscr{M})$ for short.

In our case we have

$T(\mathcal{M}_1)=T(\mathcal{P})=T(\mathcal{A})$ conjoined with the semantic assumptions I_1,
$T(\mathcal{M}_2)=T(\mathcal{R})=T(\mathcal{A})$ conjoined with the semantic assumptions I_2.

$\mathcal{M}_1$ and $\mathcal{M}_2$ above are but two members of an unlimited population of models of $\mathcal{A}$. And they are all *full models* in the sense that they are obtained by interpreting every constituent of the abstract primitive base $\mathcal{A}$. We might also have built a family of partial models resulting from a *partial interpretation* of $\mathcal{A}$. It would be the family of all the structures in which the nature of S but not the one of $\preccurlyeq$ is specified. (On the other hand it would be impossible to specify the order relation without at the same time fixing the nature of the elements of S.) In short, there are degrees of abstraction or, conversely, of semantic commitment. This notion is made more precise by

DEFINITION 6.1 Let $T(\mathcal{A})$ be an abstract theory with a primitive base $\mathcal{A}=\langle A_1, A_2, \ldots, A_n\rangle$ consisting of n nonlogical constants. Furthermore let $\mathcal{M}=\langle\mu(A_1), \mu(A_2), \ldots, \mu(A_n)\rangle$ be the value of an interpretation μ at $\mathcal{A}$. Finally, assume that μ does not effect a mere permutation (reshuffling) of the coordinates of $\mathcal{A}$. Then $\mathcal{M}$ is of *syntactic rank n*, *semantic rank* $m\leqslant n$ and *degree of abstraction* $\alpha=(n-m)/n=_{df} m$ of the interpreted primitives $\mu(A_i)$ differ from the corresponding abstract primitives A_i.

DEFINITION 6.2 $\mathcal{M}=\langle\mu(A_1), \mu(A_2), \ldots, \mu(A_n)\rangle$ is a *model* (or a *full model*) of $\mathcal{A}=_{df}$ the degree of abstraction of $\mathcal{M}$ is $\alpha=0$. On the other hand, if $0<\alpha<1$, $\mathcal{M}$ is a *partial model* of $\mathcal{A}$.

Instead of the degree of abstraction α we might have defined the *degree of interpretation* $\beta=1-\alpha=m/n$. This would have had the advantage that it would not have involved the ambiguous term 'abstraction', which we use in its semantic acceptation not in the epistemological one of remoteness from sense experience. The concept of degree of interpretation will reappear in the theory of factual interpretation (Sec. 3.4).

We close this subsection with a couple of historical notes. The idea of a partially interpreted calculus, usually credited to Carnap (1939), goes back to Boole and was used by Whitehead (1898, pp. 10–11) in his campaign for the independence of algebra with respect to arithmetic. And the notion of a partial model introduced by Definition 2 must not be taken for the concept of semimodel introduced by Kemeny (1956): a semimodel involves full interpretation and differs from a model in that it does not involve validity in some structure.

2.3. *Intensional Models and Extensional Models*

We distinguish two kinds of mathematical interpretation and consequently two sorts of model: extensional and intensional. Equivalently: a model can be characterized either extensionally or intensionally. (Recall that our use of 'intensional' is the traditional one not the one current in modal logic.) An *extensional interpretation* assigns to every predicate in the abstract theory its extension in some field. For example, a binary relation is interpreted as the set of ordered pairs that stand in the given relation. On the other hand an *intensional interpretation* maps the abstract primitives into more specific mathematical objects that need not be set theoretic objects. For example, in the class interpretation of lattice theory considered in Table 6.2, Sec. 2.1, the lattice operations (meet and join) are assigned the class intersection and the class union respectively, and these operations are in turn characterized by the axioms of the algebra of classes.

More precisely, let T be an abstract theory formalized to the extent that all of its specific primitives can be identified and arranged in a sequence

$$\mathscr{A}=\langle A_1, A_2, \dots, A_n, \dots\rangle.$$

A possible *intensional model* of $\mathscr{A}$ is a structure

$$\mathscr{M}(\mathscr{A})=\langle \mu(A_1), \mu(A_2), \dots, \mu(A_n), \dots\rangle,$$

the coordinates of which are definite mathematical objects with the same logical properties as their arguments in $\mathscr{A}$: an individual constant in $\mathscr{A}$ (e.g., the unit element in an algebra) is assigned an individual in $\mathscr{M}(\mathscr{A})$; a class in $\mathscr{A}$ is matched with a class in $\mathscr{M}(\mathscr{A})$; an m-ary relation in $\mathscr{A}$ is assigned an m-ary relation in $\mathscr{M}(\mathscr{A})$; a function in $\mathscr{A}$ is paired to a function in $\mathscr{M}(\mathscr{A})$, and so on. For example, in Table 6.2 we had the propositional interpretation of $\mathscr{L}=\langle S, \preccurlyeq, \wedge, \vee\rangle$, to which it assigned the model $\mathscr{P} = \langle P, \vdash, \&, \vee \rangle$. Only the first coordinate of this quadruple is a set.

On the other hand a possible *extensional model* of an abstract structure $\mathscr{A}$ is produced by (*a*) introducing a nonempty domain of individuals D (the domain of the model) and (*b*) interpreting every coordinate of $\mathscr{A}$ as either a member of D or a set of m-tuples of members of D. In particular, an individual constant of $\mathscr{A}$ is matched with a member of D, an abstract unary predicate in $\mathscr{A}$ is interpreted as a subset of D, and every abstract m-ary predicate of $\mathscr{A}$ is assigned a subset of D^m. Every coordinate of a model is now a mathematical object with a definite set theoretic status: see Figure 6.1. In particular, the image of an abstract predicate is the extension of the latter – not, as sometimes claimed, the meaning of the predicate.

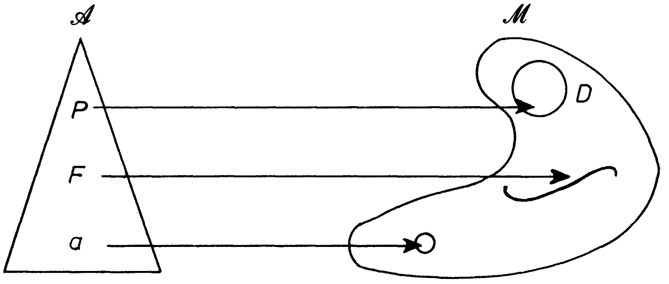


Fig. 6.1. An extensional interpretation maps the constituents of an abstract structure $\mathscr{A}$ into set theoretic objects built exclusively with the domain D of the model.

The peculiarities of the two kinds of model are exhibited in Table 6.3 overleaf.

2.4. *Insufficiency of Extensional Models*

The discipline that studies extensional models (in the sense of Sec. 2.3) is called *model theory*. This is a sizable and growing chapter of metamathematics and it may be regarded as covering most of the semantics of logic and mathematics. The concern of model theory is "the mutual relations between sentences for formalized theories and mathematical systems in which these sentences hold" (Tarski, 1954, p. 572). For example, model theory investigates the relations between abstract Boolean algebra and its models. In particular, model theory can characterize the whole lot of models of a given abstract structure and can study the morphisms between such models.

Model theory is not only concerned with models *per se* but is also interested in the *use* of models to solve certain syntactic problems about any mathematical theory, whether abstract or specific. Indeed, model theory is the most powerful tool available to investigate questions of consistency, concept independence, definability, independence of axioms, provability, categoricity, etc. On this count model theory is relevant not only to pure mathematics but also to the foundations of science and to exact philosophy.

TABLE 6.3

Intensional model and extensional model

Abstract primitive	Intensional object	Extensional object
Individual constant a	Individual $\mu(a)$	$\mu(a) \in D$
Unary predicate P	Attribute $\mu(P)$	$\mathscr{E}(P) \subseteq D$
m-ary predicate P^m	m-ary attribute $\mu(P^m)$	$\mathscr{E}(P^m) \subseteq D^m$
m-ary operation or function F^m	m-ary operation or function $\mu(F^m)$	$\mu(F^m): D^m \to D$

However, as conceived heretofore model theory is limited to extensional models and is therefore of restricted use even for purely mathematical purposes. Firstly, extensional models are not easy to come by: except in trivial cases sets are not given extensionally, i.e., by displaying their memberships, but are determined by some predicate or other. That is, normally a set is given by some law or rule which is not in turn resolvable in set theoretic terms. (Thus the fact that the *general* notion of a function can be partially elucidated as a set of ordered n-tuples does not entail that every *special* function can be so given. For example the logarithmic function is not given by a table of logarithms – the extensionalist ideal – but by certain formulas, such as $\ulcorner log(xy) = log\,x + log\,y \urcorner$, with $x, y \in R^+$.) In mathematics, just as in science, extensions are ultimately determined by senses. Secondly, even if it were possible to construct every model or example in purely set theoretic terms, intensional models could be dispensed with only provided we adopted the principle that coextensives are identical. But, as we say in Ch. 4, Sec. 1.2, this is a false dogma. It is particularly misleading with reference to factual science, where "descriptive interpretations" are essential (Carnap, 1958, p. 173). Therefore the claim (Suppes, 1961, 1967, 1969; Przelecki, 1969) that model theory can take care of the semantics of science is as unjustified as the identification of a model of an abstract structure ("formalized language") with "the real world" (Beth, 1962) or even with "a fragment of reality" (Przelecki, 1969).

Model theory does not tackle any of the problems peculiar to the semantics of factual science for the following reasons:

(i) The overwhelming majority of the mathematical theories used in factual science are not abstract but interpreted (within mathematics). Thus there is no way of reinterpreting a differential equation within mathematics: its degree of abstraction is nil. Now, model theory has little if anything to say about such theories – e.g., the theory of complex variables, the theory of integral equations, or differential geometry. Only abstract theories, such as group theory, or the general theory of topological spaces, pose model theoretic problems such as "Does this interpretation of the primitives yield a model?", "Are all the models of the given structure isomorphic to one another?", or "Can we prove a representation theorem for this theory?".

(ii) The models occurring in "intuitive" (nonformalized) mathematics and in science are mostly intensional models, i.e., they are "defined" by properties and laws rather than extensionally. On the other hand the models studied by model theory are extensional, hence incapable of discerning intensional differences unless the latter are accompanied by extensional differences. Applied mathematics and science cannot dismiss such differences, particularly since coextensive predicates may be characterized by different law statements, whence they must be counted as distinct.

(iii) As practised in formalized mathematics, which is the object of model theory, axiomatics involves *de*interpretation. For example, the abstract theory of natural numbers is formulated in such a way that the very concept of a natural number is not explicitly included in it, precisely in order to allow for alternative interpretations. A possible axiomatization of this abstract theory boils down to the following set of postulates:

A1 $x' \neq 0.$

A2 $x' = y' \Rightarrow x = y.$

A3 $(P0 \,\&\, (Px \Rightarrow Px')) \Rightarrow (y)\, Py.$

One recognizes here the kernel of the five Dedekind-Peano axioms. But the above formulas are satisfied in models other than the standard number theoretic one. In order to have the above postulates describe the essential properties of the natural numbers, suitable interpretation assumptions must be joined to them. Interpretation is thus external to formal axiomatics as opposed to the axiomatics found in "concrete" or "intuitive" mathematics and in science. (For an elucidation of the differences between formal axiomatics and *inhaltliche Axiomatik* see Hilbert-Bernays, 1968, I, Sec. 1.) In particular, scientific axiom systems must contain interpretation assumptions, as emphasized by Carnap (Carnap, 1939, 1958). Otherwise we would not know what the theory is about nor, consequently, how to apply and test it.

(iv) Because science concerns the external world, scientific theories must involve not only mathematical interpretations but also factual ones, i.e. construct-fact correspondences. These correspondences fall outside the scope of model theory, which is concerned with mapping structures into structures. The semantic assumptions in factual science correlate definite mathematical structures with real systems – and a real system is not a mathematical object. (The fashionable identification of models with possible worlds has suggested the view that the actual world is just a possible model. This new fangled version of the Platonic allegory of the cavern overlooks a couple of details. One is that, while a model is a harmless unpolluted construct, the world is not the work of a mathematician. Another is that, while a formula may or may not be satisfied in a model, the natural laws are inherent in the real world. A third is that, while every single model is fully characterized, no chunk of reality, however minute, is known exhaustively.) Furthermore the semantic assumptions in factual science are refutable hypotheses (Ch. 3). For example, improved accuracy in measurement showed that Yukawa's theory was not about μ-mesons, as originally conjectured, but about π-mesons. In contrast, the rules of assignment (of extensions) given by an extensional model may be regarded as holding analytically provided analyticity is construed in a permissive fashion (Kemeny, 1956).

In sum, model theory does not help us to elucidate the semantic peculiarities of factual science. The semantics of science takes off where model theory leaves: see Figure 6.2.

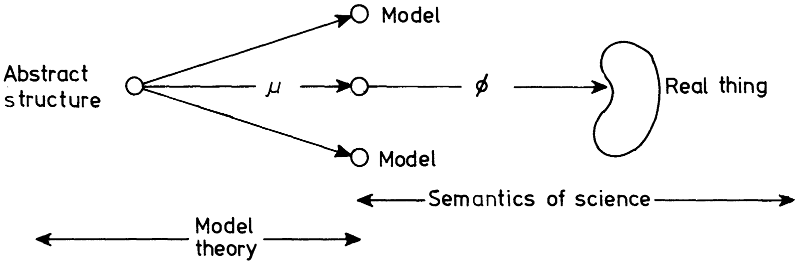


Fig. 6.2. From abstraction to reality through model (or conversely).

Let us now turn to the problem of factual interpretation – the map ϕ that model theory does not deal with.

3. FACTUAL INTERPRETATION

3.1. *The Need for Factual Interpretation in Science*

Every contemporary scientific theory worth the name has a mathematical formalism. This formalism consists of a set of mathematical theories the symbolism of which is interpreted by tacit or explicit designation rules pairing symbols to constructs. The overwhelming majority of such formalisms are not abstract theories but theories of models in the sense of Sec. 2.2. Probability theory and the theory of Hilbert spaces are theories of this kind and the two are components of the mathematical formalism of quantum mechanics. Whatever additional interpretation any such formalism be assigned, is nonmathematical – hence it cannot be described exclusively in mathematical terms. Thus a physical geometry will consist of a mathematical geometry together with semantic assumptions pairing constructs to things or properties of things.

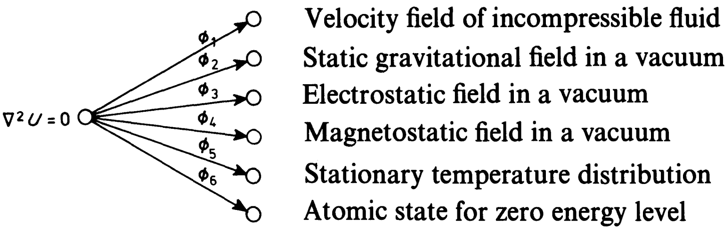
However, many mathematicians see no need for such additional interpretations. Thus a distinguished mathematical physicist postulates that the isotropic subspaces of R^n are light rays, whence ray optics is just the theory of such spaces (Jost, 1965, p. 18). Another distinguished mathematician proposes the following definition: "A lever is a system consisting of a plane π, a straight line ℓ in that plane, called a beam, a point O on that line called the fulcrum, etc." (Freudenthal, 1971, p. 316). Mind: in the two quotations certain mathematical objects are not said to *represent* certain physical objects but are *identified* with them. Finally an eminent philosopher-scientist has forcefully defended the slogan "To axiomatize a theory is to define a set-theoretical predicate", i.e., "one that can be defined within set theory in a completely formal way" (Suppes, 1967). If this is how scientific theories are to be reconstructed then it is obvious that (*a*) logic applies to physical objects such as dynamical systems: "as far as dynamical systems are concepts (a lever, a solar system) they allow for logical relations" (Freudenthal, 1971, p. 321); and (*b*) "there is no theoretical way of drawing a sharp distinction between a piece of pure mathematics and a piece of theoretical science" (Suppes, *op. cit.*). The view, held by all three authors, that a scientific theory consists of its mathematical formalism alone, may be regarded as an updated version of the Pythagorean philosophy and may be called *semantic formalism* – or *unsemantics* for short.

Most theoretical scientists are not semantic formalists: they hold with Einstein (1936) that a scientific theory has a content that overflows its mathematical formalism and thereby throws the theory at the mercy of facts. Suppes himself acts on this nonformalist conviction when expounding scientific theories. Thus he formulates his theoretical model of individual decision in the following fashion (Suppes, 1969, p. 148). "We shall call an ordered triple $\mathcal{S} = \langle S, C, D \rangle$ an *individual decision* situation when S and C are sets and D is a set of functions mapping S into C. The intended interpretation is:

S = set of states of nature,
C = set of consequences,
D = set of decisions or actions."

These interpretation assumptions are not included in the axiomatic definition of an individual decision situation but are juxtaposed to the latter: nevertheless they are not forgotten although they certainly are not set theoretic constructs.

The typical theoretical scientist will not say that a factual item f (thing, property, state, event, process) *is* a mathematical object m but rather that *m represents f*. He knows that one and the same mathematical object (set, function, space, equation, etc.) may be assumed to represent different factual items in different theories. For example, the Laplace equation occurs in at least the following capacities:


$\nabla^2 U = 0$
ϕ_1
ϕ_2
ϕ_3
ϕ_4
ϕ_5
ϕ_6
Velocity field of incompressible fluid
Static gravitational field in a vacuum
Electrostatic field in a vacuum
Magnetostatic field in a vacuum
Stationary temperature distribution
Atomic state for zero energy level

Since scientists meet the same functions and equations over and over again in different chapters and associated with different factual contents (senses and referents), they know that a scientific theory has a factual content which is not exhausted by its formalism. Most scientists realize that whatever can be read *out* of a mathematical formalism is what we have more or less unwittingly read *into* it. They differ only as to the nature of this content and the way it should be assigned. Thus while most scientists seem to favor a realist but sloppy semantics, those who do take pains to spell out the interpretation assumptions do it often in operationist terms. Only a few hold the magic view that a mathematical formalism yields its own interpretation (Everett, 1957; DeWitt, 1970). To help settle these problems let us analyze a couple of examples.

3.2. *How Interpretations are Assigned and What They Accomplish*

Let us exhibit three examples of factual interpretation with a view to finding out what it adds to the mathematical formalism.

Theory 1: Rat race

Primitives: S, Δ.

Axiom 1 Δ is a binary associative operation in the set S: i.e., for any $x, y, z \in S$, $(x\Delta y)\,\Delta z = x\Delta\,(y\Delta z)$.

Axiom 2 $\phi(S)$ = collection of rats.

Axiom 3 $\phi(\Delta)$ = joining a race, i.e., $\phi(x\Delta y) = x$ joins y (in this order) in a race to some goal.

The mathematical formalism boils down to Axiom 1, which states that

$\langle S, \Delta \rangle$ is a semigroup. The other two postulates are semantic assumptions. By virtue of the latter, the former becomes the law statement ϕ (A1) = ⌜First to enter race wins⌝ (or ⌜Late comers are eliminated⌝).

Without A2 and A3 no such "reading" (interpretation) of A1 would be explicit and unambiguous. The first semantic assumption identifies the referents of the theory as rats. The second stipulates that Δ represents race joining. Axiom 1 determines the mathematical sense of Δ, but the full sense of Δ is given by all three axioms.

Theory 2: Switching circuits

This theory contains half a dozen semantic assumptions defining an interpretation function ϕ that relates certain formulas to elements of circuitry of a kind. This function is a one to one mapping $\phi : B \to N$ from the set B of Boolean functions of a certain type onto the set N of two terminal series-parallel electrical networks. The domain of ϕ is a construct while its range is an aggregate of pieces of hardware: ϕ is a factual interpretation function. Given any Boolean form b in B, ϕ locates a possible network n in N such that $\phi(b) = n$, i.e. such that *b represents n*. And conversely: given a possible network n, its Boolean image will be $b = \phi^{-1}(n)$, where ϕ^{-1} is the inverse of ϕ. *Example*: Figure 6.3.

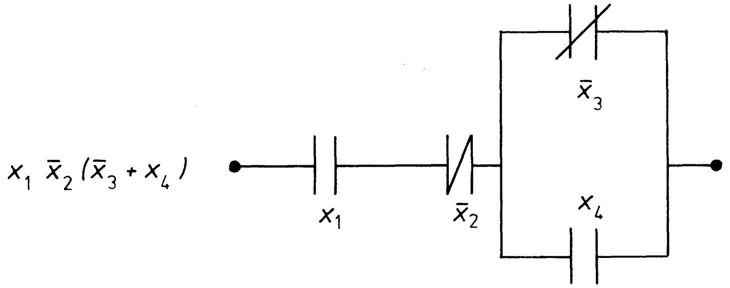
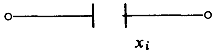


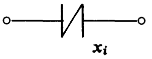
Fig. 6.3. A representation of switching circuits by Boolean forms.

The specific semantic assumptions determining ϕ are (Harrison, 1965, p. 79)

A1 $\phi(0) = \circ$ (Equivalently: $0 \hat{=} \circ$)
A2 $\phi(1) = \circ\!\!-\!\!-\!\!-\!\!\circ$ (Equivalently: $1 \hat{=} \circ\!\!-\!\!-\!\!-\!\!\circ$)

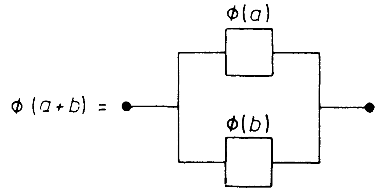
A3 For any variable x_i, $\phi(x_i) =$ 

(Equivalently: $x_i \hat{=}$ a normally open contact)

A4 For any variable x_i, $\phi(\bar{x}_i =$ 

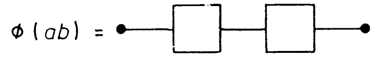
(Equivalently: $\bar{x}_i \hat{=}$ a normally closed contact)

A5 For any Boolean forms a, b in B:



(Equivalently: $a+b \hat{=}$ two terminal parallel circuit.)

A6 For any Boolean forms a, b in B:



(Equivalently: $ab \hat{=}$ two terminal series circuit.)

Here again the semantic assumptions determine both the reference class and the way the constructs represent some of the features of their referents.

Theory 3: Assembly theory

The preceding considerations apply not only to scientific theories but also to theories in scientific or mathematical metaphysics, such as assembly theory (Bunge, 1971b). This theory is concerned with the basic modes of assembly or composition of systems apart from their specific properties. It may be regarded as ring theory (a stalwart member of abstract algebra) together with the following semantic assumptions

A1 $\phi(S) =$ the set of all systems
A2 $\phi(0) =$ the null system
A3 $\phi(+) =$ system juxtaposition or joining
A4 $\phi(\cdot) =$ system interpenetration or superposition

By virtue of these semantic assumptions every formula in ring theory becomes a metaphysical statement. For example, for any x, y, z in S,

$\phi(x(y+z)=xy+xz)=$ The superposition of system x with the outcome of the juxtaposition of systems y and z equals the juxtaposition of the systems (x superposed to y) and (x superposed to z).

We now generalize the preceding considerations by defining a theoretical factual construct as a mathematical construct together with a factual interpretation map. More precisely, we adopt

DEFINITION 6.3 A construct c will be said to be a *theoretical factual construct* iff

(i) c belongs to a theory and

(ii) $c=\langle m, \phi\rangle$, where m is a mathematical construct and ϕ is an interpretation map such that $\phi(m)$ is a factual item (thing, property, or event) or a collection of factual items.

Example The ordered pair $\langle M, \phi\rangle$ is the concept of mass in particle mechanics iff $M: P \rightarrow R^+$ is an additive function and

(i) $\phi(P)=$ Particles,

(ii) $\phi(M(x))=$ Inertia of x for every $x \in P$,

(iii) M occurs in the equations of motion of particle mechanics multiplying the particle acceleration.

To conclude we collect the lessons learned from our analysis:

(i) The nonsemantic axiom(s) of a theory determine(s) the mathematical sense of its primitives;

(ii) the semantic axiom(s) determine(s) the referents and sketch(es) the full factual sense of the primitives and of the nonsemantical axiom(s);

(iii) the sense and reference of the derived constructs of a theory are determined by its axioms.

In sum, the sense and reference of a theory are determined jointly by all of its axioms. Equivalently: the significance of the symbolism ("language") of a theory is given by all of its axioms taken together.

3.3. *The Factual Interpretation Maps*

The first three examples discussed in the last subsection are extremely simple, hence atypical: they involved a single interpretation. In fact, in each of them the interpretation function ϕ mapped an abstract structure $\mathscr{A}$ into a factual domain $\mathscr{F}$. Thus in the first case $\mathscr{A}=\langle S, \Delta\rangle$, and $\mathscr{F}$

consisted of a collection of racing rats. There was no intermediary mathematical model such as, say, the ring of integers or Euclidean geometry. In brief, we had

$$\mathscr{A} \overset{\phi}{\mapsto} \mathscr{F}. \tag{1}$$

Moreover in the second case ϕ was no less than an isomorphism between the set B of constructs and the collection N of things. Besides, in this case as well as in the case of assembly theory ϕ was a morphism of addition and of multiplication. Such simplicity is exceptional in science.

In most scientific theories the domain of ϕ is not an abstract structure but a model of such. In other words, the mathematical formalism of the typical scientific theory is a *theory of a model.* And this theory is seldom found ready-made in a mathematical rack: the theory is usually built by enriching an interpreted mathematical theory (or rather a motley collection of fragments of interpreted mathematical theories) with some specific assumptions not found in mathematics. For example, a classical field theory is obtained by putting the following components together: (*a*) the theory of differentiable manifolds, (*b*) a set of specific formulas, chiefly the field equations, boundary conditions, and constraints, and (*c*) a set of semantic assumptions.

In these cases we have two successive and dovetailing interpretations: μ and ϕ, the former from an abstract structure $\mathscr{A}$ to a model $\mathscr{M}$, the latter from $\mathscr{M}$ to a factual domain $\mathscr{F}$:

$$\mathscr{A} \overset{\mu}{\mapsto} \mathscr{M}, \qquad \mathscr{M} \overset{\phi}{\mapsto} \mathscr{F}. \tag{2}$$

Hence, with a qualification to be mentioned shortly, we may regard the factual interpretation of an abstract structure as the composition of the two mappings, i.e. $\phi \circ \mu : \mathscr{A} \to \mathscr{F}$. (Bunge, 1972b.) It is true that the abstract structure is seldom if ever dug out when analyzing a typical scientific theory: one usually starts from a model. However, we do not get the full semantical picture unless we uncover that deeper layer.

In point of fact only a portion $\mathscr{M}_0$ of a mathematical model is usually assigned a factual interpretation. For example, not every vector decomposition, or series expansion, or integral representation, nor even every solution of a differential equation is always assigned a factual counterpart. A part of the mathematical formalism of a factual theory is usually either idle or plays a purely syntactical role. (For example, in Theory 3 of Sec. 3.2 the unit of the ring is assigned no special interpretation.) Consequently ϕ is normally a *partial* function from $\mathscr{M}$ to $\mathscr{F}$. Equivalently: ϕ is a total function on a *sub*set $\mathscr{M}_0$ of $\mathscr{M}$. We indicate this by writing

$$[\mathscr{M}]\overset{\phi}{\mapsto}\mathscr{F} \tag{3}$$

and drawing Figure 6.4.

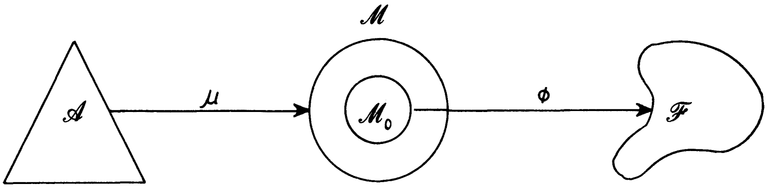


Fig. 6.4. Normally a factual interpretation ϕ maps only a part $\mathscr{M}_0$ of a model $\mathscr{M}$ of an abstract structure $\mathscr{A}$ into a factual domain $\mathscr{F}$.

In the case of the switching circuits theory discussed in Sec. 3.2 the interpretation map ϕ was one to one and consequently ϕ^{-1} had an inverse. That is, two constructs were the same (different) just in case their factual images were identical (different). We warned already that this is the exception rather than the rule: usually ϕ does not discern among equivalent systems. Actually this is so even in the case of the theory of switching circuits, which does not distinguish among circuits constructed with different materials and having different sizes as long as they are topologically equivalent. That is, ϕ is really a mapping of $\mathscr{M}_0 \subset \mathscr{M}$ into a *family of equivalence classes of concrete systems.* In other words, the range of ϕ is not a factual domain $\mathscr{F}$ but the quotient of $\mathscr{F}$ by an equivalence relation $\sim$, i.e. $\mathscr{F}/\sim$. This equivalence relation is defined in a tacit fashion by the very theory T in question, namely thus: Two factual items are *equivalent with respect to* T iff T does not discern between them, i.e. iff it represents them by the same constructs. In brief, instead of (3) we usually have

$$\phi : [\mathscr{M}] \to \mathscr{F}/\sim, \tag{4}$$

where $\mathscr{M}$ is a collection of mathematical models and $\mathscr{F}$ a set of factual domains.

In sum, we distinguish four different factual interpretation maps ϕ:

Abstract theory → Factual systems (1)
Theory of a model → Factual systems (2)
Part of a theory of a model → Factual systems (3)
Part of a theory of a model → Equivalence classes of factual systems. (4)

In either case we adopt the following

DEFINITION 6.4 If $\mathscr{M}$ is a mathematical model and ϕ a factual interpretation map, then the structure $\mathscr{M}_\phi = \langle \mathscr{M}, \phi \rangle$ is called a *factual model.*

DEFINITION 6.5 If $\mathscr{M}_\phi$ is a factual model, then a theory $T(\mathscr{M}_\phi)$ of such a model is called a *factual theory.*

The preceding elucidations suffice to settle certain much debated questions such as whether gravitation theory is reducible to geometry, as is often held with reference to Einstein's general relativity. If the semantic assumptions of the theory are not stated explicitly, then of course one has, strictly speaking, only a mathematical formalism not a factual theory. However, the context makes it usually clear what things (material systems and fields) the theory is about and which of the properties (e.g. gravitational interaction) it represents. That is, the formalism is treated as a full fledged factual theory. Moreover, and this is a strictly syntactic point, if gravitation theory were just a geometrical theory, then it would contain no formulas other than those of Riemannian geometry – but in fact it adds to the latter its own field equations and equations of motion.

3.4. *Factual Interpretation: Full and Partial*

We saw in the preceding subsection that the mathematical formalism of a theory is likely to contain components without representatives in the real world. For example, not every Fourier analysis will represent the spectral decomposition of a wave packet. Conversely, not every trait of a factual system is likely to be represented by any theory of it. Thus half of the wave-like solutions to Maxwell's field equations are usually discarded because they would have to be interpreted as representing waves coming from the future. (They are often called 'unphysical' or 'physically meaningless' solutions.) On the other hand the same equations fail to account for the photon structure of a light beam. Briefly: the theory has redundant constructs and also leaves some factual items dangling in the midst of reality. This is quite general: we may assume that no $\mathscr{M}$ is isomorphic with $\phi(\mathscr{M})=\mathscr{F}$ except perhaps in some limited respects. Usually $\mathscr{M}$ contains elements with no images in $\mathscr{F}$ and, conversely, $\mathscr{F}$ has elements not accounted for by $\mathscr{M}$. See Fig. 6.5.

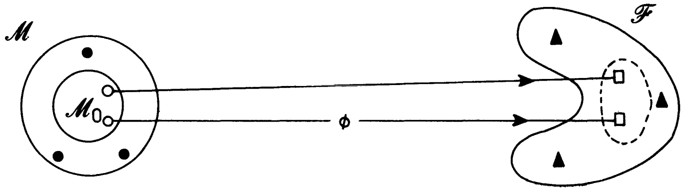


Fig. 6.5. Black dots = redundant constructs. Black triangles = orphan entities.

From a semantic point of view the best scientific theory referring to a given factual field is the one with the fewest black spots and leaving the fewest black triangles. This holds for the formulas (e.g., equations) and their constituents (e.g., parameters in the equations). Rich theories are likely to contain idle formulas but on the other hand are bound to contain few if any uninterpreted parameters, while shallow theories abound in such parameters. (A *fundamental theory* is often defined as containing no constants other than universal constants.) While a redundant formula can be isolated and immobilized, the factually uninterpreted parameters cannot be disposed of short of replacing the theory. Such parameters are of the essence of phenomenological (or black box) theories as well as of the hypotheses that cover the available data and little else. Since those parameters can be varied almost *ad libitum* to conform to the data, the corresponding theory is a docile data recipient with only a weak explanatory power (Bunge, 1963b, 1964, 1967a). The more detailed a "picture" of reality a theory supplies, the more heavily interpreted it is: the less specific, the less semantically committed.

The degree of semantical commitment of a scientific theory can be quantified with the help of the Definition 1, in Sec. 2.2, of abstraction degree. We presently adapt it to a factually interpreted model and a factually interpreted theory as characterized by Definitions 4 and 5 in Sec. 3.3:

DEFINITION 6.6 Let $\mathcal{M}=\langle M_1, M_2,..., M_n\rangle$ be a mathematical model of some abstract structure and let ϕ be a factual interpretation of $\mathcal{M}$. If $m \leqslant n$ of the interpreted concepts $\phi(M_i)$ for $1 \leqslant i \leqslant n$ differ from the corresponding mathematical primitives M_i, then the factual model $\mathcal{M}_\phi = \langle \mathcal{M}, \phi\rangle$ and any theory $T(\mathcal{M}_\phi)$ of it are said to possess *a degree of interpretation* $\beta = m/n$.

DEFINITION 6.7. Let $\mathcal{M}_\phi = \langle \mathcal{M}, \phi\rangle$ be a factual model and $T(\mathcal{M}_\phi)$ be a theory of $\mathcal{M}_\phi$. Then

(i) $\mathcal{M}_\phi$ and $T(\mathcal{M}_\phi)$ are said to be *fully interpreted* in factual terms iff $\beta = 1$.

(ii) $\mathcal{M}_\phi$ and $T(\mathcal{M}_\phi)$ are said to be *partially interpreted* in factual terms iff $0 < \beta < 1$.

Example 1 Probability theory is based on two primitives: the sample space S and the probability measure P. If neither is assigned a factual interpretation, $\beta = 0$ and the theory remains in pure mathematics. If the elements of S are interpreted as states of a system or as events of some kind, e.g. learning a certain task, then $\beta = \frac{1}{2}$. Most stochastic learning theories are of this kind, i.e. semantically half committed. Finally, if also P is interpreted, then $\beta = 1$. For example, if $s \in S$ is assigned an event of a certain kind, then $P(s)$, which is some number in the real number interval [0, 1], might be interpreted as the tendency or disposition for that event s to occur. This would complete the interpretation of the stochastic theory concerned – without of course ensuring its truth. (On the other hand identifying $P(s)$ with the relative frequency of s would not count as an interpretation of probability but as an estimate of probability values.)

Example 2 Let $\mathcal{M} = \langle S, F, G, k\rangle$ where S is a set, F and G real valued functions on S, and k a positive real. So far this is a specific structure or model. Now introduce an interpretation map ϕ such that

$$\phi(S) = \text{set of bodies}, \ \phi(F) = \text{mass}, \ \phi(G) = \text{volume}, \ \phi(k) = k.$$

We get a factual model $\langle \mathcal{M}, \phi \rangle$ with degree of interpretation $\beta = \frac{3}{4}$. And if a different map ϕ' is chosen, one that assigns to every coordinate in $\mathcal{M}$ a factual item, then β jumps to 1.

3.5. *Generic Partially Interpreted Theories*

A *generic factual theory* is one concerned with a genus rather than a species of factual items – e.g., with bodies, or organisms, or societies. When adjoined specific assumptions it can become a specific theory prepared to deal with certain traits of some species of systems – say fluids, or flat worms, or industrial societies. In short, the reference class of a generic theory in factual science is some genus or other – not necessarily a natural genus. And what it represents are some pervasive features of its referents.

There are two kinds of factual generic theory: fully interpreted and partially interpreted in factual terms. The giant classical theories of science, such as classical mechanics and the theory of evolution, are fully interpreted generic theories: they are concerned with whole families of species and all their basic concepts are assigned a factual interpretation. In addition to these theories there are generic factual theories with a low degree of interpretation. This lack of a firm semantical commitment makes them readily portable from one field of research to another. The first outstanding specimen of this kind of theory was Lagrangian dynamics, which started out as a chapter of mechanics, then spread over almost all of physics (cf. Bunge, 1957, 1967b) and has now found its way into the general theory of systems of any kind (White and Tauber, 1969). Newer members of the species are information theory, the mathematical theory of machines, and general network theory. All of these theories are factual in the sense that they refer to real systems though not to any definite species of such. And, far from representing specific properties, they represent only very general features. Consequently they are useful guides in either of the following situations: (*a*) absence of detailed knowledge of the system; (*b*) a detailed knowledge may be available but only certain outstanding features compatible with a number of possible mechanisms are of interest, and (*c*) a unified treatment of a number of research subjects is intended – for instance in order to bring out their common formal traits.

To get a feel of the semantic peculiarities of such generic semi-interpreted theories in factual science, as well as of the methodological problems they raise, we shall take a look at the Rashevsky-Turing theory of morphogenesis (Rosen, 1970, I, Ch. VII). This theory states, in a nutshell, that any initially homogeneous or amorphous system that reaches an unstable state may evolve towards a final state of inhomogeneity (e.g., polarity) under the action of slight external perturbations. The state variables of this theory are left factually uninterpreted. Only the independent variable is interpreted, namely as time. Moreover, the theory is kinetic rather than dynamic, in the sense that it assumes no specific forces or interactions responsible for the process: any force will do as long as it is compatible with the equations of change of state. In sum, the Rashevsky-Turing theory of morphogenesis is a morphological theory: a theory of the genesis of differentiation or form in nearly any complex system. It is richer than a black box theory in that it accounts for certain changes in the interior of the box, but is equally noncommittal regarding the nature of the components and their interactions.

A Rashevsky-Turing system is defined as any system satisfying the assumptions of the theory, whatever its actual physics and chemistry may be. In other words, the Rashevsky-Turing theory has a number of possible interpretations. It is not just that it concerns a whole class of systems – every general theory does. It concerns a family of classes, i.e. a genus. As soon as the state variables of the theory are specified, i.e. as soon as they are assumed to represent definite properties, interactions, etc., one species in the genus is singled out for reference. In short, upon interpreting the state variables of the theory, the family of species is restricted to a single species of morphogenetic systems.

This semantic difference between a partially interpreted and a fully interpreted theory is of importance for methodology. Since a generic morphogenesis theory specifies neither the substratum nor the forces acting in it, no definite predictions can be computed with the formulas of the theory. Hence such a theory is untestable in the usual way. Partially interpreted theories call for a revision of the conventional methodology of science. In fact such theories are tested vicariously, namely by testing some of the specific theories obtained upon specifying (interpreting) the state variables as definite properties of a system of a definite kind (Bunge, 1973a, Ch. 2).

3.6. *Principles of Factual Interpretation*

Any given mathematical construct may be assigned a number of different factual interpretations. Therefore there is often uncertainty, and occasionally spirited controversy, as to which interpretation is the best. Consequently it is desirable to have a battery of explicitly formulated criteria for an admissible factual interpretation – if not to facilitate the interpretation task at least to ease rational argument about it. We propose the following conditions that a sound interpretation of a mathematical construct in factual terms should satisfy: the interpretation should concern a reasonably safe mathematical construct; it should not originate inconsistencies; it should be strict, i.e. adjusted to the mathematical formalism; it should be literal not figurative; it should be factual rather than empirical; it should be full not partial; and it should aim at the truth. Let us spell out these conditions.

(i) *Factual interpretations should bear on mathematically sound formalisms.* If the mathematical skeleton is ambiguous or inconsistent, no amount of cunning interpretation will turn it into a reasonable factual theory. This seems self-evident, yet some highly refined scientific theories, such a quantum electrodynamics, fail to satisfy the condition: they contain ambiguous expressions (e.g., integrals whose value depend on the mode of computation) and inconsistencies (e.g., the electric charge, assumed to be finite when occurring in an equation of motion, turns out to be infinite in derived formulas). Hence the interpretation of such theories must be regarded as insecure. And instead of trying to save the sick formalism by a semantic *tour de force* one should try alternative formalisms. But before anyone attempts to bell this cat the dogma that quantum electrodynamics is perfect will have to be shaken.

(ii) *Factual interpretations should introduce no inconsistencies.* There is danger of inconsistency whenever a construct is assigned different factual correlates – i.e. if the theory contains more than one interpretation map. However, this is sometimes necessary and need not lead to any inconsistencies. For example, a neuropsychological theory may contain variables that are assigned both a neurological and a psychological interpretation. Thus in Grossberg's theory of learning networks every vertex function is interpreted both as a stimulus trace and as an average membrane potential (Grossberg, 1969). These happen to be mutually compatible interpretations of one and the same mathematical construct. On the other hand the standard formulations of quantum mechanics contain multiple interpretations that do lead to contradiction – as when the 'Δx' occurring in Heisenberg's inequalities is interpreted as both the average scatter of the *particle* position and the width of the *wave* packet, and perhaps also as the physicist's *uncertainty* concerning the exact particle position. (See Bunge, 1973b.)

(iii) *Factual interpretations should be strict not adventitious.* A factual interpretation should match the structure of the construct concerned: it should pour no more content than the construct can possibly hold. For example, if a hamiltonian contains only variables referring to a given system (e.g., a molecule) then it should not be taxed with representing both the system and an unspecified measuring device – let alone the experimenter's mind. In general, the value of a function should not be interpreted as concerning more referents than arguments. If it be assumed that a formula α concerns a certain fact f, then α should contain at least one variable x such that $\phi(x)=f$. Otherwise it must be concluded that the interpretation is adventitious: that it has no mathematical leg to stand on (Bunge, 1969).

(iv) *Factual interpretation should be literal not metaphorical.* In mathematics the concept of analogy can be elucidated in an exact way, namely as homomorphism, and can thus be kept under control. Outside mathematics analogy wears too many faces, all of them blurred and tinged by subjectivity: one man's similarity is another's dissimilarity. Metaphor can be a pragmatic asset: it may have heuristic value and may also be of some use in teaching but it can be terribly misleading precisely for being highly subjective. For this reason it does not belong in scientific theory – *pace* a rather fashionable view (Black, 1962; Hesse, 1965). The aim of a new scientific theory is not to win neophobic followers but to account for things with characteristics of their own – which traits the metaphor is bound to hide, as the gist of metaphor is to pass the new for old. A scientific theory must involve literal interpretations only – no *as ifs*. This need was first realized at the turn of the century when

Maxwell's electromagnetic theory was freed from any mechanical associations, and is now acutely felt in relation with quantum mechanics. Indeed the classical analogies of position and momentum, and particle and wave, though probably inevitable in the early stages, have introduced inconsistencies and have blocked the understanding of the theory as an original creation referring to *sui generis* things (Bunge, 1967c). In short: analogy belongs in the scaffolding not in the building. (Metaphor is dead but it won't lie down.)

(v) *Scientific theories should be interpreted by reference to facts not test procedures.* For example, acidity is indicated by, but not interpretable as, a color change in litmus paper; and physiological stress is not interpretable as organ enlargement as revealed by autopsy. If on the other hand an analysis of observation (or of measurement) is held to be the key to the factual interpretation of a theory, then (*a*) meaning is being mistaken for testability and (*b*) the scope of the theory is restricted to situations under experimental control. Which is what happens with the standard or Copenhagen interpretation of quantum mechanics. The result is not only confusion but also inconsistency, as exemplified by Bohr's thesis that the theory, though nonclassical, is based on classical physics (i.e., presupposes the latter) just because the end results of measurements are describable in classical terms. If operationism is given up then the quantum theories can be interpreted in their own revolutionary terms – as demanded, albeit timidly, by Wheeler (1957) and Everett (1957) – as well as in strictly objective terms (Bunge, 1967b).

(vi) *Factual interpretations should be global not spotty.* Not isolated formulas but whole formalisms should be assigned a factual interpretation if the risks of inconsistency and irrelevance are to be avoided. Taken in isolation any formula can be interpreted in a number of ways: taken together with other formulas of the same conceptual system, the number of interpretations is decreased because the number of conditions to be satisfied is increased. For example, Shannon's formula for the quantity of information (cf. Ch. 4, Sec. 3.2) looks like Boltzmann's formula for the entropy and is therefore often interpreted as the entropy of the system. However, this metaphoric interpretation is quite arbitrary, as the information theoretic "entropy" is not related to any thermodynamic function such as energy, temperature, pressure, or volume. Consequently it is not entitled to be called 'entropy' – nor is entropy entitled to pass for quantity of information. Just as mathematical interpretation is constrained by the requirement that it yield formulas satisfiable in some model, so factual interpretation should produce formulas reasonably true to fact – in particular it should ensue in statements representing laws. In other words, factual interpretation is not a matter of convention nor even one of mathematical validity: it depends upon the actual structure of the world. Which borders on the next condition.

(vii) *Factual interpretation should maximize truth.* The semantic assumptions of a scientific theory should contribute to producing a maximally true theory. As with the previous conditions, this one is easier legislated than lived up to. Since factual truth depends on both the mathematical formalism and the semantic assumptions, the goal of maximal truth can only be attained by a mutual adjustment of these two components. The test of correctness of semantic assumptions is of course the truth of the theory as a whole. But we can never check the entire theory and we should not expect it to be completely true. Hence even a strong confirmation of the theory provides no final assurance that the semantic assumptions are right. And if the tests are unfavorable then we may blame either the formalism or the semantic assumptions and attempt to reform either. Whatever the outcome of the tests we cannot be sure about the adequacy of the interpretation. We must take risks and be prepared to lose. In sum, interpretation is just as tentative as formalism – and both are prior to tests. By the same token interpretations can be changed in the interest of truth. If a theory fails to pass some tests for truth it need not be rejected *in toto*: some of it may be salvaged by partially modifying its formalism or its interpretation or both. In any case interpretation is prior to truth valuation and should maximize truth value.

This last condition leads us to the next point – the confusion between interpreting and stipulating truth conditions.

3.7. *Factual Interpretation and Truth*

Interpretation maps constructs either into further constructs (the case of μ) or into facts (the case of ϕ). In either case interpretation is prior to truth valuation: the latter depends upon the former. Thus consider the abstract formula $\ulcorner$For every x and z there is at least one y such that $x \circ y = z\urcorner$. Unless we interpret the individual variables and the operation we cannot even ask whether the formula holds. A formula in abstract mathematics holds or fails to hold *under some interpretation* (or in some model) or other. Of course we are chiefly interested in interpretations conducive to truth, so that an interpretation that fails to satisfy this condition will be given up. Likewise in factual science interpretation comes before truth valuation even though an unfavorable outcome of the latter may force us to reinterpret the given mathematical formalism. In sum, both in mathematics and in factual science *only interpreted formulas can be tested for truth*, and only such tests allow us to assign truth values. In a nutshell, the process looks like this:

Formulation → *Interpretation* → *Testing* → *Truth valuation.*

The same can be seen from another angle: interpretation and truth valuation are utterly different functions. Let us confine ourselves to factual interpretation ϕ and to the assignment $\mathcal{V}$ of degrees of truth of fact. (But similar considerations hold for mathematical interpretation and assignment of formal truth.) For one thing ϕ applies not only to full formulas but also to their nonlogical constituents, whereas $\mathcal{V}$ is a (partial) function on statements. For another, whereas the range of ϕ is a factual domain, the range of $\mathcal{V}$ is a set of truth values, e.g., 0 and 1. Consequently *to give the semantics of a scientific theory does not involve giving truth conditions*, let alone truth values: it only requires *specifying the interpretation map* ϕ.

However, according to a widespread view interpreting involves or even consists in giving truth conditions, perhaps even truth values. Thus Carnap: "By a *semantical system* (or interpreted system) we understand a system of rules, formulated in a metalanguage and referring to an object language, of such a kind that the rules determine a *truth-condition* for every sentence of the object language, i.e. a sufficient and necessary condition for its truth. In this way the sentences are *interpreted* by the rules, i.e. made understandable, because to understand a sentence, to know what is asserted by it, is the same as to know under what conditions it would be true. To formulate it in still another way: the rules determine the *meaning* or *sense* of the sentence" (Carnap, 1942, p. 23; see also p. 203). And, a quarter century later, Davidson (1967, p. 310): "to give truth conditions is a way of giving the meaning of a sentence".

This influential view is a version of the verification doctrine of meaning suggested by Frege and proposed by the operationists, Wittgenstein, and the Vienna Circle. It is so confusing that the reasons for rejecting it bear some hammering. First, although the doctrine looks plausible for the propositional calculus, where the sense of the connectives may be said to be given by their truth tables, the view fails for predicate logic. Here both the individual variables and the predicate variables have to be interpreted in a manner that is independent of truth, as shown in any standard logic textbook (e.g. Mendelson, 1963, pp. 49ff.; Shoenfield, 1967, pp. 61ff.; Suppes, 1957, pp. 64ff.). Second, before we set out to find out the truth value of a formula we must know what it "says" about what: fancy placing truth conditions on an uninterpreted formula. Third, truth depends upon interpretation, not the other way around. Thus $\ulcorner(\exists x)\, Gx\urcorner$ is true for $\phi(G)=$ Gipsy, and false for (or "under") $\phi(G)=$ Ghost. Fourth, except for idealist philosophers the assignment of degrees of factual truth is not a matter of semantics but of observation and scientific inference. Semantics cannot even devise truth conditions for scientific hypotheses and theories: this is a matter for methodology. Thus consider a theoretical statement of the form

$$t=\ulcorner P(s,u)=n\urcorner$$

evaluated in the light of a piece of empirical evidence in the form

$$e=\ulcorner\text{Average of measured values of } P(s,u)=n'\pm\varepsilon\urcorner,$$

where P is some property of a system s, n is the calculated value and n' the measured value (both in units u), while ε is the experimental error. Then a "truth condition" for t that is universally agreed on (without the assistance of the available semantical theories) is this:

$$t \textit{ is true relative to } e\textit{, to within } \varepsilon\textit{, iff } |n-n'|\leqslant\varepsilon.$$

The actual value of the experimental error ε will depend on the state of the experimental art: it is no business of semantics. (For the empirical assessment of truth values see Bunge, 1963a, pp. 127ff; and Bunge, 1967a, II, pp. 301ff.)

In short, interpretation and truth are related but not the way operationist semantics has it. Truth depends upon interpretation, which in turn should be subject to revision in the light of the tests for truth. A formula will hold or fail to hold (exactly or approximately), under a given interpretation, whereas alternative interpretations may render the formula meaningless or utterly false. This holds for mathematics as well as for science. So much for one of the worst muddles in the history of philosophy.

3.8. *Interpretation and Exactification*

Factual constructs come with different degrees of exactness and clarity. The most exact and clear of all are those belonging to a theory, i.e., the theoretical factual constructs. According to Def. 3 (Sec. 3.2) a construct c of this kind is a mathematical construct m together with a factual interpretation ϕ, i.e., $c = \langle m, \phi \rangle$. Consequently to *exactify* a factual construct consists in either revealing or assigning its formal component m. And to *elucidate* a factual construct consists in either revealing or assigning its semantic component ϕ. If we disclose the form or the content of a construct we engage in analysis; if we assign either of them we build or reconstruct a fragment of one of the theories housing the construct of interest.

In principle every *bona fide* scientific construct can be both exactified and elucidated, namely by either incorporating or expanding it into a theory or, if it belongs already to some theory, by analyzing or reconstructing the latter. Even initially obscure concepts can be exactified and elucidated. A good example is the one of disposition, tendency, propensity, or bent, which pervades both factual science and metaphysics. This intuitive notion can be analyzed into two distinct concepts: those of *causal propensity* and *chance propensity* (cf. Bunge, 1974b). An instance of the former is solubility: dissolution is an outcome of the mixing of the soluble substance with a proper solvent under suitable circumstances. Whenever these conditions are met dissolution occurs. Not so chance propensity, as exemplified by the emission of light by an atom or the learning of an item by an animal: even if the necessary conditions are met, the event occurs only with a certain probability – i.e., there seem to be no conditions both necessary and sufficient for the event to happen. Let us focus on this second concept of tendency, which is by far the more baffling and probably the most fundamental of the two.

The intuitive or pretheoretical concept of chance propensity is exactified by means of the mathematical concept of probability. And any specific concept of chance propensity is elucidated by incorporating it into a factual theory; for example, every concept of learning disposition or ability is elucidated by the corresponding stochastic learning theory. Exactification, though essential, is not sufficient to turn a specific concept of chance propensity into a semantically precise concept, because the mathematical theory of probability is not committed to any particular factual interpretation. We must specify also the interpretations of the arguments and values of the probability function. (Recall Example 1 in Sec. 3.4.) This may be done in the following way. Let the pretheoretical idea be that of the tendency or ability of a system σ of kind Σ to make a transition from an initial state A to a final state B. (For example, Σ could be a strain of albino rats, A the state of ignorance concerning the proper way to run a T maze, and B some stage in the learning process.) The explicans of that relatively obscure notion of ability is the ordered pair $\langle Pr(B \mid A), \phi \rangle$, where $Pr(B \mid A)$ is the conditional probability of B given A, and ϕ the interpretation map defined by the following value assignments:

$$\phi(\sigma) = \text{System of kind } \Sigma \tag{1}$$

$$\phi(A) = \text{Initial state of } \sigma \tag{2}$$

$$\phi(B) = \text{Final state of } \sigma \tag{3}$$

$$\phi(Pr(B \mid A)) = \text{Strength of the propensity for } \sigma \text{ to jump from } A \text{ to } B. \tag{4}$$

In other words, the coarse or presystematic idea of the tendency for σ to go from A to B is given a refined (exact and lucid) expression by the theoretical factual construct $\langle Pr(B \mid A), \phi \rangle$, which belongs to some theory concerning certain traits of the systems of kind Σ – a theory whose mathematical formalism includes some fragment of the mathematical theory of probability. While the latter is in charge of the *exactification* of the concept of chance propensity, the interpretation assumptions (1) to (4) provide an *elucidation* (or semantic clarification) of it. Let us insist that the interpretation assumptions are not part of the exactification procedure but external to it. If we were to regard them as belonging to the exactification process we would fall into a circle: we would be explaining propensity as propensity.

The preceding considerations solve one of the problems posed by the so-called propensity interpretation of probability championed by Popper (1959). The problem is to answer the charge that nothing is gained and much is lost by interpreting the clear concept of probability in terms of the obscure concept of propensity. Our answer is this: There is nothing wrong with adopting at the same time the propensity *interpretation* of probability, i.e.,

$$\phi(\text{Probability}) = \text{Propensity},$$

and the probability *exactification* of propensity, i.e.,

$$\varepsilon(\text{Propensity}) = \text{Probability},$$

as long as the two are not mixed up. Whereas interpretation assigns a factual content to a definite mathematical construct, exactification transforms an inexact construct into a definite mathematical object. Shorter: whereas exactification bears on presystematic concepts, interpretation enriches exact concepts.

Moreover just as the probability exactification of chance propensity is consistent with the propensity interpretation of probability, both are compatible with the *frequency estimate* (evaluation) of probability values. For example, in game theory the probability that a player chooses a given strategy may be interpreted as his propensity to adopt such a choice (cf. Rapoport, 1966); and this value can be estimated by observing the actual frequency of such an event. What is not possible is to give a *frequency interpretation* of probability. For one thing, probability and frequency are different functions: the latter is defined, for every sampling procedure, on a finite subset of the whole probability space; and the range of the frequency function is not the whole real interval [0, 1] but the collection of fractions in it. (Cf. Bunge, 1969.) Consequently if probabilities were *interpreted* as frequencies, the typical theorems of the calculus of probability, such as the laws of large numbers, could not even be stated, as they concern precisely *differences* between probabilities and frequencies. (For further criticisms of the frequency theories of probability see Fréchet (1939) and Suppes (1967, Ch. 3.)) Interpretation is a strictly conceptual operation not to be confused with numerical estimation, in particular measurement. Such a confusion amounts to mistaking semantics for pragmatics – something that deserves a separate section.

4. PRAGMATIC ASPECTS

4.1. *Pragmatic Interpretation*

School teachers find it effective to clarify mathematical and scientific ideas with reference to human operations. Thus $\ulcorner 3+2=5 \urcorner$ can be made clear and plausible by fingering, and Archimedes' law of the lever can be literally felt when riding a see-saw. These are examples of *pragmatic interpretation* or interpretation in terms of human actions. In Sec. 2 we mentioned some pragmatic interpretations of the propositional calculus. Table 6.4 exhibits pragmatic interpretations of a few typical formulas of the predicate calculus. The pattern is this: every construct in a set C is assigned one item in a set H of human actions. In short, $\pi: C \rightarrow H$. In other words, π is a rule or instruction for handling a construct with a definite means and a definite goal.

TABLE 6.4

Examples of pragmatic interpretation of typical formulas

Construct	Semantic interpretation	Pragmatic interpretation
Pa	The individual a has the property P.	Someone has proved or observed that a is a P.
$(\exists x)_U Px$	There is at least one object in U with the property P.	At least one object in the observed collection $T \subseteq U$ has been found or may be found to be a P.
$(x)_U Px$	All objects in U have the property P.	Every object in the observed collection $T \subseteq U$ has been found or may be found to be a P.
$A \vdash B$	A entails B.	B is provable from A.
$f(x)=y$	The f-ness of x equals y.	The result of determining (computing or measuring) f at x is (nearly) y.
$P_n(x)=0$	The nth degree polynomial P_n in x equals 0.	Find the values of x that annul the nth degree polynomial P_n in x.

There can be no quarrel with pragmatic interpretations when employed as didactic crutches – particularly if being reassured that the crutches will be dropped in due time. Nor should it be objected to translating formulas into instructions or commands for purposes of computer processing, lab checking, or action – particularly if the formulas are allowed to retain a content of their own independent of the way they are used or tested. What is inconvenient is to have to walk on crutches all life long; even worse, to be under the delusion of being a computer or of being chained to some measuring device. In other words, what is objectionable is to mistake a construct for a pragmatic interpretation of it. It is even worse to dignify this confusion with the name of a philosophy – say operationism, operative logic, or mathematical intuitionism. In short, while pragmatic interpretations are occasionally valid and useful (though always restricted to a minute subset of the collection of constructs), pragmatist semantics is untenable.

Most pragmatic interpretations are *adventitious* in the sense of Sec. 3.6. Indeed, in most cases they do not match the structure of the formula they bear on, as they refer to individuals (e.g., observers) and actions (e.g., measurements) not represented in the formula by any variables. For example, the orthodox interpretation of an eigenvalue α_k of a quantum-mechanical operator A_{op} reads: "α_k is a possible result of measuring the property represented by A_{op}". This interpretation is adventitious because neither A_{op} nor α_k (nor the corresponding eigenfunction) contain any variables capable of representing the measuring device (which one?) or the experimenter (who?). (Cf. Ch. 3, Sec. 4.3.)

If we were to outlaw all adventitious interpretations, few pragmatic interpretations would remain. As long as we are aware of the narrowness and arbitrariness that adventitious interpretation may land us on, we may afford to adopt a wider concept of interpretation validity. We propose the following conditions for regarding a pragmatic interpretation of a formula as *valid* (Bunge, 1969):

(i) There should exist a scientific theory containing the formula and assigning to it a semantic (mathematical or factual) interpretation. In other words, the formula to be interpreted (*a*) must be available to begin with and (*b*) must have a fairly definite content independent of the many ways it can be manipulated. (Imagine rushing to read a new scientific theory in operational terms before finding out what the sense and the reference of the theory are.)

(ii) Sufficient information, theoretical and empirical, must be at hand to justify as well as to carry out the operations called for or described by the pragmatic interpretation. If the construct on which the interpretation bears happens to represent an unobservable entity or property, as is so often the case with science, then additional hypotheses or theories linking the unobservable to observable items will be needed. (I.e., objectifiers or indicators will be needed, and these will usually rope in further theories.) Otherwise the proposing of a pragmatic interpretation would be a game like casting horoscopes or interpreting dreams. In other words valid pragmatic interpretation, even when adventitious, is a matter of law not convention: there should exist a lawful relation between the referent of the construct and the human action the rule prescribes. In sum, pragmatic interpretation should be grounded.

Pragmatic interpretation occurs in experimental science and in technology. The experimentalist may well read $\ulcorner y=f(x) \urcorner$ as "To infer y measure x" provided f is defined and the semantic interpretation tells him what these symbols stand for. Likewise the engineer may read the same formula as, say, "To get output y apply input x" – as long as the underlying theory supplies him with the sense and reference of the formula and provided experiment encourages him in taking the assumed functional relation as close enough to the truth. Pragmatic interpretations such as these are valid albeit adventitious: they rely on the formula concerned as well as on its semantic interpretation. Similarly with every other pragmatic interpretation: if valid it will rest on a previously assigned semantic interpretation. First get to know, then apply your knowledge.

There is no room for pragmatic interpretation in scientific theories. A theoretical formula refers to some concrete system (cell, society, or what have you) not to the way that very formula is to be tested or applied. Even the sciences of action, such as operations research and political science, regard their referents as objects. Consequently their formulas are first assigned a semantic interpretation, then they may be put to use as rules of procedure. We had to emphasize the dependence of pragmatic interpretation upon semantic interpretation because of the strong human tendency, called *anthropomorphism*, to read everything in terms of human feelings and actions. We must rid semantics of any association with such a tendency if we want it to account for the objectivity of science.

4.2. *The Interpretation Process*

Interpretations do not come out of the blue and, once proposed, they need not stay. Viewed historically, interpretation is a process. In some cases the formalism of a theory and its interpretation evolve hand in hand. In others the embryo is an intuitive idea in search of a formalism: this may have been the case with Newtonian mechanics, Maxwellian electrodynamics, and Einsteinian gravitation theory. Finally the converse process, namely the building of a formalism in search of an interpretation, can also happen: as a matter of fact this seems to have been the case, to a large extent, with quantum mechanics (Dirac, 1942; Heisenberg, 1955). Consequently there are no hard and fast rules for "discovering" the semantic assumptions of a scientific theory: some investigators proceed in one way, others differently. It is up to the psychology of science, not to semantics or even methodology, to discover what makes researchers tick, in particular what makes them guess that a given formula should be interpreted in a certain way.

Moreover no interpretation is likely to be final. Every theory in a process of growth suffers both mathematical and semantical adjustments. Even classical theories are still undergoing changes of both kinds (cf. Truesdell and Toupin, 1960). In particular the new interpretation may be at variance with the original intentions of the first theorist. The psychologist and the historian of science may wish to ask him what interpretation he had in mind but won't be able to discover all the possible unintended interpretations – if only because most of them will never occur to anyone. In any case the concept of intention, that occurs in the usual phrase 'intended interpretation', is a psychological one and therefore beyond the reach of semantics. Whether a particular outcome, semantic or other, was originally intended, is a psychological and historical problem. Therefore it is misleading to define interpretation as an *intended model* of a formalized language (Kemeny, 1956). For the same reason it is unsatisfactory to mention the *intended interpretation* of a formalism without stating it explicitly; and, once so stated, it is no longer intended. The semantic assumptions of a theory, whether they are the originally intended (or standard) ones or not, should be formulated as explicitly as the remaining assumptions if we are to have objectivity and the possibility of rational argument.

The need for arguing about interpretation matters is not always felt. It seems most acute in highly developed fields – but there it is often repressed. Every theoretical biologist knows that it is far easier to interpret the solution to a problem in mathematical biology than it is to formulate the problem. The opposite situation is the rule in physics, where it is far easier to formulate a problem and even to perform the computational tasks it calls for, than to find an adequate interpretation of the solution. Why this difference? In biology there are hardly any comprehensive theories supplying a general framework for the formulation of problems. Except in choice areas such as biophysics and genetics, nearly every problem has got to be treated separately, often leaning more heavily on physics or chemistry than on biology. Usually theories have to be built from scratch, sometimes inaugurating whole new branches of biology in the process. As a compensation the goal is more modest: the variables involved are fewer, they are often better understood and are frequently linked in simpler ways than the variables occurring in theoretical physics and chemistry. We can expect however that, as biology grows in depth, it will pose as many and as hard interpretation problems as physical theories do at present.

Regretfully, rational argument about the semantic assumptions of scientific theories is sometimes discouraged or even hushed up. Even complex theories like quantum mechanics and quantum electrodynamics, ridden as they are with unsettled interpretation problems, are often regarded as unproblematic and all discussion about their semantic assumption as a waste of time (Rosenfeld, 1961). There are several possible motives but not a single reason for adopting such a dogmatic and unhistoric stand. One is the longing for certainty. Another is the belief that foundations problems are settled by popular philosophical discourse rather than by digging out the axiomatic foundations of the theory concerned. A third possible cause is a faulty semantics of science, one holding that all that really counts in a scientific theory is its mathematical formalism. If this were true, then the producing of any new formula or any new set of numbers would be a valuable contribution to scientific knowledge, whereas the proposing of a more cogent interpretation of a theory would be insignificant. This attitude is widespread among scientists who must devote most of their time to solving difficult computation problems, e.g., with the help of perturbation theory. They take the basic equations for granted and consider themselves lucky if once in a blue moon they can find solutions in closed form – which are the ones best suited to interpretation. Since they have little time left to ponder over the interpretation of their very starting points, they have no patience with anyone's telling them that interpretation is always problematic, hence deserving of closer scrutiny. But of course the belief is mistaken. Since a theory in science is a formalism together with an interpretation, a change in the latter produces a new theory. Besides, some interpretations deserve being reformed because they are wrong. Hence disputes over interpretation matters are as important as arguments over mathematical points. What is true, and unfortunate, is that the standards of argumentation over semantical problems are so much lower than the standards of mathematical argument. It behooves the philosopher to upgrade those standards by building a semantical theory competent to deal with live science.

5. CONCLUDING REMARKS

In our view, since meaning is sense together with reference, a meaning assignment is an assignment of both sense and reference. Such an assignment involves an interpretation of the symbols concerned and eventually also an interpretation of the constructs designated by the symbols – as when the letter 'N' is first read as the cardinality of a set, then as the population of a group of organisms. However, we do not regard interpretation as meaning assignment. One reason for not identifying these two concepts is that, while interpretation may bear either on signs (e.g., predicate letters) or on constructs (e.g., functions), we construe meaning as a property of constructs only (see Ch. 7). Another reason is that not every interpretation assigns significance: some interpretations result in nonsignificant expressions. For example, if the predicate letter P in '5 is P' is interpreted as 'painful', a nonsignificant sentence results. Interpretation, though necessary, is insufficient to guarantee significance. Significance derives from meaning, which is in turn a conceptual matter.

Even assuming that we grasp the general concept of meaning we may not know how to go about assigning or finding out specific meanings. Placing the given construct (concept or proposition) in a given context (e.g., a theory) is surely necessary to that end since meaning is contextual, but may not suffice. Thus the axioms of an abstract theory, such as Boolean algebra, determine the (mathematical) sense of the theory but, fortunately, fail to specify any of the possible referents of the theory. In other words, the sets involved in the abstract theories are abstract: they consist of faceless individuals. It is only when adjoining an interpretation that definite individuals are characterized by the theory. And only semantic assumptions of the factual type indicate that such individuals are factual items. The semantic assumptions of a factual theory determine, *inter alia*, the reference class of the theory and thus contribute to sharpening its factual sense. In terms of the symbolism or language of a scientific theory: the axioms of such a theory (all the axioms) determine both the sense and the reference the symbolism points to. Or, as we shall say, they jointly determine the meaning of the factual theory. This, in a nutshell, is the theory of meaning to be developed in the next chapter.

CHAPTER 7

MEANING

We are now in a position to face the chief scandal of semantics. We shall start by distinguishing the significance of a sign from the meaning of the construct it designates. We shall then proceed to formulate and discuss our view that meaning is no more and no less than sense together with reference. If either component changes then a changed meaning, i.e., a different construct ensues. Since our previous investigations have shown us how to disclose sense and spot referents, we shall be in a position to compare meanings. In particular we shall be able to ascertain whether two given constructs have the same meanings so that their respective symbols will be synonymous. Finally we shall discuss some of the difficulties that have hampered the clarification of the meaning concept.

1. BABEL

Although the concept of meaning has been the subject of active inquiry since Socrates, and at the center of analytic philosophy for half a century, it is still far from clear. To be sure there have been heaps of brilliant meaning analyses and plenty of talk about the theory of meaning and the theory of reference, particularly about the virtues such theories ought to exhibit. But in fact no theory proper has been offered that does justice to the two aspects of meaning traditionally discerned, namely sense (or connotation) and reference (or denotation). And none of the existing theories of meaning, not even those proposed by philosophers of science, has been of any help whatever in performing a semantic analysis of a piece of live science or in teaching scientists to speak sensibly about the meaning of their own creations. No wonder many physicists still claim that the meaning (not just the worth) of a theoretical item is determined by observation procedures. No wonder chemists are fond of saying that each triplet of bases means (not just specifies or determines) a particular aminoacid. No wonder geneticists sometimes hold that mutations can produce meaningless sequences – rather than biologically dysfunctional proteins. In sum half a century of talk about meaning has been useless to scientists and has, if anything, increased confusion among philosophers. It has resulted in a Babel.

The views on meaning come in several degrees of formal sophistication but their gist can be conveyed in plain words. The most influential among the contemporary views are displayed schematically in the following list.

1 *Psychologism:* meaning is either thought or intention or understanding.
2 *Pragmatism:* meaning is usage.
3 *Operationism:* meaning is operation (computation or measurement).
4 *Verificationism:* meaning is truth condition.
5 *Epistemic view:* meaning is information.
6 *Nihilist view:* there are no meanings.
7 *Referential view:* meaning is the thing referred to.
8 *Intensional view:* meaning is either intension or content.
9 *Dualist view:* meaning has two dimensions – intension and extension.
10 *Synthetic view:* meaning is composed of sense and reference.

The first six views are characteristically modern. Psychologism is the oldest of them but was not made academically respectable until Brentano expounded it one century ago. It was espoused at one time by Russell (1919b). The second, linguistic pragmatism (and in particular dictionarism), was propounded by the later Wittgenstein. The third, operationism, may be traced back to Peirce and Dingler (1907) and has been in the air natural scientists have been breathing ever since. The fourth, verificationism, may be traced back to Frege and was one of the battle cries of the Vienna Circle. The fifth, informationism, was born in the 1950's. And the sixth, nihilism, is a cry of despair over the failure of all the previous opinions. We have criticized and rejected all six in Chs. 2, 4 and 5.

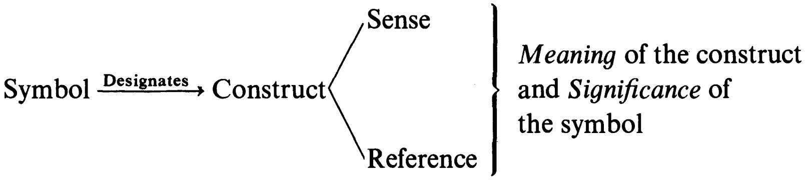
The remaining four views have much longer roots and are far sounder. The referential doctrine goes back to the medieval nominalists, particularly Ockham and Buridan, and their modern heirs, especially Hobbes. The intensionalist doctrine has been a constant component of idealism probably since Plato and was particularly vivid in Leibniz and Bolzano (1837). The dualist doctrine was sketched in the Port-Royal *Logique*

(1662) and was later revived, but also obscured, first by Frege (1891, 1892), then by Lewis (1944, 1951).

(We may as well list the weaknesses of Frege's influential doctrine of meaning in order to forestall any confusion between it and the next or synthetic view. To begin with, the great Frege did not employ a consistent terminology: for example, he often exchanged *Bedeutung* (our "reference") and *Bezeichnung* (designation). He did not distinguish clearly between reference and extension. He often construed *Sinn* (sense) in a psychologistic fashion, namely as the thought expressed by a sentence. He identified the *Bedeutung* of a statement with its truth value and ominously assigned truth conditions the job of determining senses. Finally Frege had no semantic theory to speak of: he never advanced beyond some unsystematic remarks – albeit often illuminating and always provocative. Frege's importance for semantics seems to lie in that (*a*) he stressed a number of distinctions – notably between symbol and construct, concept and extension, and sense and truth, and (*b*) he called the attention of others, notably Russell and Carnap, to semantic problems. For a very different evaluation see Dummett (1973, Ch. 19).)

Finally the tenth view – the synthetic doctrine – seems to have a medieval pedigree too. It was rescued from oblivion and popularized by J. S. Mill: his *System of Logic* (1843) was so influential that his distinction between the connotation and the denotation of a term has been incorporated into common speech. This view appealed to common sense and, since Mill was a champion of positivism, his revindication of connotation or sense, against the grain of nominalist referentialism, was above the suspicion of Platonism. Our own view is but an elaboration of Mill's and of Williams' (1937), which incorporated Mill's without its positivism. We have called it *synthetic* for the following reasons: (i) it applies not only to terms and other linguistic expressions but also (and primarily) to their conceptual designata; (ii) while it distinguishes sense from reference, it combines them into a single idea with a definite mathematical status – namely the sense-reference ordered pair; (iii) far from being a stray view it is an outcome of our theories of sense and of reference expounded in previous chapters. Let us peek at it before going into details.

To begin with we stipulate the kinds of object that can possess meaning. They are certain symbols and all constructs. In order to ward off confusion we shall call these two possibilities by different names: we shall say that (some) *symbols signify* and that (all) *constructs mean.* Furthermore we shall take meaning as primary and shall define it by its two components: sense and reference. And we shall construe signification, a property of certain signs, as the composition of designation and meaning. Graphically:



For example we shall say that the term ‘man’ designates the (or rather *a*) concept of man, the sense of which is given by anthropology, its reference class being the whole bunch of humans. These two constitute also the significance or vicarious meaning of the word ‘man’. On the other hand any sign that fails to designate a construct will be assigned an empty significance. For example a punctuation mark stands for no construct and is therefore devoid of significance. In sum if a sign is significant it is so vicariously, namely via some construct. This construal of ‘meaning’ avoids both nominalism and the variety of hylemorphism consisting in endowing mere marks with semantical properties. And it does not commit us to Platonism, for we do not adopt the ontological hypothesis that constructs have an independent being. Furthermore our view formalizes the opinion of those linguists who hold that a word has two semantic functions: one is to denote, the other is to encapsulate a whole system of generalizations and associations (Luria, 1961).

2. THE SYNTHETIC VIEW

2.1. *Meaning as Sense cum Reference*

In the previous section we agreed to regard meaning as a property of constructs (concepts, propositions, or theories). Now we propose to analyze the meaning of a construct as its sense together with its reference, as depicted in Figure 7.1 and exemplified in Table 7.1.

The stipulation that sense and reference are to be taken as the two components of meaning is literal not metaphorical. Calling $\mathscr{S}(c)$ the sense and $\mathscr{R}(c)$ the reference class of a construct c, our proposal can be symbolized as

$$\mathcal{M}(c) = \langle \mathcal{S}(c), \mathcal{R}(c) \rangle .$$

Now, as we saw in Chapters 4 and 5, $\mathcal{S}$ maps constructs into sets of constructs, i.e. $\mathcal{S}: C \rightarrow \mathcal{P}(C)$. And, according to Chapter 2, $\mathcal{R}$ maps constructs into sets of objects of any kind, i.e. $\mathcal{R}: C \rightarrow \mathcal{P}(\Omega)$, where $\mathcal{P}(\Omega)$ is

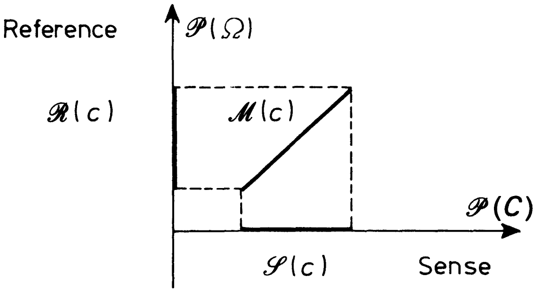


Fig. 7.1. Sense and reference as the components of meaning.

TABLE 7.1

The two components of meaning of the rat race theory in Chap. 6, Sec. 3.2

Construct	Meaning components	
	Sense	Reference
S	Set	Rats
Δ	Joining a rat race	Rats

the set of all subsets of the collection Ω of objects. The pairs $\langle \mathcal{S}(c), \mathcal{R}(c) \rangle$ define uniquely a third function

$$\mathcal{M}: C \rightarrow \mathcal{P}(C) \times \mathcal{P}(\Omega)$$

such that

$$p \circ \mathcal{M} = \mathcal{S} = \text{The first projection } (= \text{component}) \text{ of } \mathcal{M}$$
$$q \circ \mathcal{M} = \mathcal{R} = \text{The second projection } (= \text{component}) \text{ of } \mathcal{M}$$

That is, the two triangles in the following diagram commute:

$$\begin{tikzcd}
& & C \arrow[dll, "\mathscr{S}"'] \arrow[d, "\mathscr{M}"] \arrow[drr, "\mathscr{R}"] & & \\
\mathscr{P}(C) & & \mathscr{P}(C)\times\mathscr{P}(\Omega) \arrow[ll, "p"'] \arrow[rr, "q"] & & \mathscr{P}(\Omega)
\end{tikzcd}$$

In other words, we lay down the following

DEFINITION 7.1 Let Ω be the universe of objects and $C\subset\Omega$ the collection of constructs. Call $\mathscr{S}:C\rightarrow\mathscr{P}(C)$ the (full) sense and $\mathscr{R}:C\rightarrow\mathscr{P}(\Omega)$ the reference functions. Then the bijection $\mathscr{M}:C\rightarrow\mathscr{P}(C)\times\mathscr{P}(\Omega)$, such that $\mathscr{M}(c)=\langle\mathscr{S}(c),\mathscr{R}(c)\rangle$ for c in C, is called the *meaning function*, and its value $\mathscr{M}(c)$ at c is called the *meaning of c*.

Remark 1 Note that the full sense is involved in the preceding – i.e., the union of the purport and the import of the construct concerned (recall Ch. 5, Sec. 5). *Remark 2* This construal of meaning does not depent on the concept of truth, as both sense and reference are prior to any assignment of truth value. Both Lewis' (Lewis, 1944, 1951) and the author's (Bunge, 1967a) previous definitions as the intension-extension pair are unacceptable for this reason. (They are not identical because Lewis construes intensions in a referential fashion.) *Remark 3* On any extensionalist semantics theology is meaningful or meaningless depending on one's religious beliefs. According to our view theological statements can be perfectly meaningful in their own contexts, which determine both their sense and their reference. Belief and disbelief must rest on the assignment of extensions, not of sense or of reference. Thus while for a theist $\mathscr{E}$(Creator)$=\mathscr{R}$(Creator)$=\{$God$\}$, for an atheist $\mathscr{R}$(Creator)$=$ $=\{$God$\}$, but $\mathscr{E}$(Creator)$=\emptyset$. Consequently if anyone wished to argue for or against a particular religion he could not count on our semantics: he would have to resort to alternative means. In particular the disbeliever won't get away with the cheap assertion that theology is nonsense. (But he may succeed in showing that some theologies are contradictory, or that all of them lack positive empirical evidence.) On the other hand the claim that existentialism and Zen Buddhism make hardly any sense stands.

We can now give an exact answer to a question left over from the days of logical positivism, viz., Are tautologies meaningful? In Ch. 2, Sec. 3.3 we saw that a tautological construct refers to anything; if universal, like $\ulcorner(x)\,(Px \vee \neg Px)\urcorner$, it refers to the totality Ω of objects. And in Ch. 5, Sec. 4, we saw that the sense of a tautological construct in a context $\mathbb{C}=\langle S, \mathbb{P}, D\rangle$ with underlying logic L equals $S \cup L$. Consequently if t is any universal tautology in L,

$$\mathscr{M}_{\mathbb{C}}(t)=\langle S \cup L, \Omega\rangle .$$

If the context in question is strictly logical, i.e. if $S=L$, then t "says" whatever L "says", and this about any objects:

$$\mathscr{M}_{L}(t)=\langle L, \Omega\rangle .$$

But of course the extralogical meaning of t in L is $\langle\emptyset, \Omega\rangle$. I.e. within logic the tautologies "say" nothing extralogical about everything. And when associated with some extralogical body of knowledge they "say" everything the latter "says" because they get stuck to every bit of it. Hence logic cannot, by itself, teach us anything about the world even when we make it speak about the world: logic is not ontology. Whatever logic succeeds in teaching us about the world it does so by associating with extralogical contexts. In conclusion tautologies are meaningful even if they do not inform us about the world. (More in Bunge, 1974a.)

Consider now the whole meaning space, i.e. the totality of values of the meaning function $\mathscr{M}$. Take the set $\mathbb{P}$ of all predicates (alternatively propositions) concerning a fixed universe of discourse $D \subset \Omega$. Further, call $\mathscr{M}(\mathbb{P})$ the totality of meanings carried by the constructs in $\mathbb{P}$. The following operations can be defined in $\mathscr{M}(\mathbb{P})$: for any p and q in $\mathbb{P}$,

Meaning addition: $\mathscr{M}(p)+\mathscr{M}(q)=\langle\mathscr{S}(p) \cup \mathscr{S}(q), \mathscr{R}(p) \cup \mathscr{R}(q)\rangle$,

Meaning product: $\mathscr{M}(p) \dot{\times} \mathscr{M}(q)=\langle\mathscr{S}(p) \cap \mathscr{S}(q), \mathscr{R}(p) \cap \mathscr{R}(q)\rangle$,

Meaning complement: $-\mathscr{M}(p)=\langle\overline{\mathscr{S}(p)}, \overline{\mathscr{R}(p)}\rangle$.

Clearly the two binary operations are associative and commutative. Moreover meanings are idempotent: $\mathscr{M}(p)+\mathscr{M}(p)=\mathscr{M}(p)$ and similarly for the product. Also, multiplication distributes on both sides over addition:

$$\mathscr{M}(p) \dot{\times} (\mathscr{M}(q)+\mathscr{M}(r))=(\mathscr{M}(p) \dot{\times} \mathscr{M}(q))+(\mathscr{M}(p) \dot{\times} \mathscr{M}(r))$$
$$(\mathscr{M}(p)+\mathscr{M}(q)) \dot{\times} \mathscr{M}(r)=(\mathscr{M}(p) \dot{\times} \mathscr{M}(r))+(\mathscr{M}(q) \dot{\times} \mathscr{M}(r)).$$

Finally, by combining complement and product we build the null or least element $\square$ of $\mathscr{M}(\mathbb{P})$, whilst combining complement and addition yields the unit or last element ▯:

Null meaning: $\mathscr{M}(p)\dot{\times}(-\mathscr{M}(p))=\langle\mathscr{S}(p)\cap\overline{\mathscr{S}(p)}, \mathscr{R}(p)\cap\overline{\mathscr{R}(p)}\rangle=$
$=\langle\emptyset,\emptyset\rangle=\square.$

Universal meaning: $\mathscr{M}(p)+(-\mathscr{M}(p))=\langle\mathscr{S}(p)\cup\overline{\mathscr{S}(p)}, \mathscr{R}(p)\cup$
$\overline{\mathscr{R}(p)}\rangle=\langle\mathscr{P}(\mathbb{P}),\mathscr{P}(D)\rangle=\text{▯}.$

The set $\mathscr{M}(\mathbb{P})$ is not closed under addition and multiplication because, as we saw in Ch. 5, in general $\mathscr{S}(p)\cup\mathscr{S}(q)$ does not equal the sense of some compound of p and q. Only intensions behave this way. Hence if we restrict sense to intension the preceding considerations prove that the structure $\langle\mathscr{M}(\mathbb{P}), \square, \text{▯}, +, \dot{\times}, -\rangle$ is a ring of idempotents with unit and zero, i.e. a Boolean ring. We shall not pursue this line but shall instead analyze two notions of meaning relation.

DEFINITION 7.2 Let $p, q\in\mathbb{P}$ be either predicates or propositions with definite meanings. Then the meaning of p is *part* of ($\preccurlyeq$) the meaning of q iff the meaning of p does not add anything to that of q:

$$p\preccurlyeq q=_{df}\mathscr{M}(p)+\mathscr{M}(q)=\mathscr{M}(q).$$

DEFINITION 7.3 Let $p, q\in\mathbb{P}$ be either predicates or propositions with definite meanings. Then p and q are *semantically unrelated* ($\urcorner\!\llcorner$) iff the product of their meanings is nil:

$$p\urcorner\!\llcorner q=_{df}\mathscr{M}(p)\dot{\times}\mathscr{M}(q)=\square.$$

COROLLARY 7.1 Let $p, q\in\mathbb{P}$. Then

(i) If $p\preccurlyeq q$ then $\mathscr{S}(p)\subseteq\mathscr{S}(q)$ and $\mathscr{R}(p)\subseteq\mathscr{R}(q)$.

(ii) If $p\urcorner\!\llcorner q$ then $\mathscr{S}(p)\cap\mathscr{S}(q)=\emptyset$ and $\mathscr{R}(p)\cap\mathscr{R}(q)=\emptyset$.

Finally consider the whole lot $\mathscr{T}$ of theories with a common meaning core – e.g. the theories in mathematical linguistics. Because of the shared meaning the following sets will be nonempty for any two members T_i, T_k of $\mathscr{T}$:

$$T_i\sqcap T_k=\{s \mid T_i\vdash s \quad\text{and}\quad T_k\vdash s\}$$
$$T_i\sqcup T_k=\{s \mid T_i\vdash s \quad\text{or}\quad T_k\vdash s\}$$

(In the notation of Tarski (1956) the first set is $T_i\cdot T_k$, the second T_i+T_k.)

The former is the infimum (*glb*) and the latter the supremum (*lub*) of $\{T_i, T_k\}$. Hence we have proved

THEOREM 7.1 The structure $\mathscr{L} = \langle \mathscr{T}, \sqcap, \sqcup \rangle$, where $\mathscr{T}$ is the set of theories with a common core, is a lattice.

Let us now move on to matters of significance.

2.2. *Significance*

Let us begin by restricting our considerations to expressions belonging to a conceptual language $\mathscr{L}$ free from ambiguity. In this case the relation $\mathscr{D}$ of designation may be construed as a function from expressions Σ^{**} of $\mathscr{L}$ into constructs (see Ch. 1, Sec. 3.2). We stipulate that significance is a property that signs acquire when they happen to designate constructs – which is the case of numerals but not of musical notes. A sign of this kind signifies its significance, which is in turn the meaning of the construct it symbolizes. Briefly, signification (the function taking significances as values) is the composition of designation and meaning. In symbols: $\mathscr{S}ig = \mathscr{M} \circ \mathscr{D}$. More explicitly,

$$\begin{tikzcd}
\Sigma^{**} \arrow[rr, "\mathscr{S}ig"] \arrow[dr, "\mathscr{D}"'] & & \mathscr{P}(C) \times \mathscr{P}(\Omega) \\
& C \arrow[ur] &
\end{tikzcd}$$

In other words, we are proposing

DEFINITION 7.4 Let Σ^{**} be the set of expressions of a conceptual language $\mathscr{L}$, $\mathscr{D}: \Sigma^{**} \to C$ the designation function, and $\mathscr{M}: C \to \mathscr{P}(C) \times \mathscr{P}(\Omega)$ the meaning function. Then the composition $\mathscr{M} \circ \mathscr{D} = \mathscr{S}ig$ is called the *signification function* of $\mathscr{L}$ and its value $\mathscr{S}ig\,(s, \mathscr{L})$ for any s in $\mathscr{L}$ is called the *significance* of s in $\mathscr{L}$.

COROLLARY 7.2 The significance of a sign belonging to a conceptual language equals the meaning of the construct designated by the sign, i.e.

If $\mathcal{D}sc$ in $\mathcal{L}$, and $\mathcal{M}(c)=\langle\mathcal{S}(c), \mathcal{R}(c)\rangle$, then $\mathcal{S}ig(s, \mathcal{L})=\langle\mathcal{S}(c), \mathcal{R}(c)\rangle$.
Proof By Definitions 1 and 4.

The first component or projection $\mathcal{S}(c)$ of $\mathcal{S}ig(s, \mathcal{L})$ may be called the *vicarious sense* of s, and the second coordinate $\mathcal{R}(c)$ its *vicarious reference*. These names convey the idea that, although signs are physical objects and therefore devoid of conceptual properties, if they proxy for constructs they indirectly acquire a meaning. This vicarious meaning is their significance. All this is expressed succinctly and exactly in the following diagram.

$$\begin{tikzcd}
& \Sigma^{**} \arrow[d, "\mathcal{D}"] \arrow[ddl, "\text{Vicarious sense}"'] \arrow[ddr, "\text{Vicarious reference}"] & \\
& C \arrow[d, "\mathcal{M}"] \arrow[dl, "\mathcal{S}"'] \arrow[dr, "\mathcal{R}"] & \\
\mathcal{P}(C) & \mathcal{P}(C)\times\mathcal{P}(\Omega) \arrow[l] \arrow[r] & \mathcal{P}(\Omega)
\end{tikzcd}$$

Table 7.2 illustrates these concepts.

TABLE 7.2

The dual significance of signs: four examples

Sign	Vicarious sense	Vicarious referents
$\circ$	Concatenation	Unspecified individuals
$+$	Addition	Numbers
V	Velocity	Physical systems
U	Utility	Objects and persons

All the previous considerations hold for a language in which each sign represents just one construct. (The converse need not hold: one and the same construct may be represented by two or more signs.) The language dependence of the signification function is made explicit in the following conventions.

DEFINITION 7.5 A sign s is *significant* in a language $\mathcal{L}$ iff s designates a construct in $\mathcal{L}$.

DEFINITION 7.6 A sign s is *nonsignificant* (or *syncategorematic*) in a language $\mathscr{L}$ iff s designates no construct in $\mathscr{L}$.

An isolated individual variable letter such as 'x' and a stray individual predicate letter such as 'F' are nonsignificant precisely because they are stray, i.e. outside any theory that may assign them definite senses. As Frege (1912) warned, it won't do to say that they have variable meanings or even indeterminate meanings. But if a sign belongs to the symbolism of a definite theory then it signifies even though its vicarious reference be indeterminate – as is the case with an abstract theory. For, in this case, the sense of the symbol is determined by the axioms of the theory in question – and sense is sufficient for meaning.

The last two definitions are relevant to the one-sign-one construct languages. These are unusual: most actual languages are ridden with ambiguities, i.e. they do not possess a designation function but rather a designation relation. For example, in informal mathematical contexts the definite integral sign may designate any of a dozen or so concepts of integral: Cauchy, Riemann, Stieltjes, Lebesgue, Schwartz, etc. The precise significance of any expression involving the integral sign will depend then on the precise interpretation assigned this ambiguous sign. (So much so that some such expressions make no sense whatsoever under certain interpretations of the snake symbol. For example, the Riemann integral of the Dirac delta is as nonsignificant as '1/0'.) In other words in most cases a language, even a mathematical one, will be the result of superposing two or more different languages equipped with one designation function each. When significance questions arise, the expert starts by analyzing the given mixture. And then he may use our previous definitions.

Note finally that a sign designating a construct with nil sense is significant. On the other hand a sign such as 'spinning and inconsistent' is insignificant because "spinning" and "inconsistent", are defined on disjoint domains. (Likewise a zero velocity is a velocity, whereas hopes have no velocity because the velocity function is not defined for them.) By contrast 'round square' is significant because both "round" and "square" are defined on the same domain, i.e. the set of plane figures. This is what makes it possible to disprove the statement that there are figures both round and square. Incidentally, extensionalism cannot cope with this fact.

2.3. *Significance Assignment*

Unlike meaning, significance is *assigned* to its bearers rather than being inherent in them. By itself, i.e. apart from a more or less precise significance assignment, a sign is just a physical object. Hence it is mistaken to ask 'What does x signify?'. One should ask instead 'What significance has x been assigned in the language $\mathscr{L}$?'. Or, in pragmatic terms, 'What are we supposed to think of or do at the sight of x?'. In other words, it is up to us whether and if so what a sign is to signify. Not so with the constructs symbolized: even though we need not conceive of them as Platonic Ideas, they must be regarded as being born with some meaning – for otherwise they are nothing. In other words, whereas symbols are conventional and therefore substitutable, the constructs they symbolize are subject to law – logical, mathematical, or scientific. Consequently the disclosure of significance rests on knowledge rather than being the subject of some "language game".

If we wish to find out the standard significance of an ordinary language expression we look it up in a dictionary. But if the expression belongs to a mathematical language or to a scientific symbolism, then one must look it up in the respective theory. In either case we come up with the vicarious sense and the vicarious reference of the sign. The two go hand in hand even if one or the other happens to be somewhat hazy. Thus when elucidating the significance of an operation symbol such as '+' we must bring in the set or sets on which the operation (not the symbol) is "defined" – and the members of this set are precisely the referents of the operation (a construct). Similarly with the elucidation of the significance of scientific terms: here too the determination of the (vicarious) referents of a symbol is part of the assignment of its (vicarious) sense. In particular, a predicate symbol will signify a property of the (vicarious) referent of the symbol. Therefore it might be thought that there is no point in keeping sense and reference as distinct components of meaning. It might be surmised that reference is a function of sense. But this is not quite so: (i) an axiom set such as Peano's, precise though it is, does not characterize completely its objects or referents but fits a number of them; (ii) a scientific theory initially intended to represent things of a certain kind may turn out to refer to a different kind of object. And even if sense were to determine reference in an unambiguous way, this would not invalidate our Definitions 1 and 4, for surely an ordered couple remains such if its second coordinate is determined by the first one – as is the case with the pair $\langle x, y\rangle$, where $y=f(x)$.

Keeping sense and reference distinct has definite advantages. In the case of mathematics it allows us to elucidate the differences in meaning among the various realizations of a given abstract formalism, namely thus. Here we have as many meanings as interpretations (or as models). The first coordinate of each such meaning consists of a fixed sense determined by the abstract theory enriched with the semantic assumptions that determine the particular interpretation. And the second coordinate of the meaning value is the domain of individuals, which varies from interpretation to interpretation. What is common to all these is of course the sense of the abstract theory. And in the case of factual science the distinction between the two meaning components is a useful reminder that one and the same thing may be represented by a number of constructs. Here, unlike the case of mathematics, the second coordinate is held fixed: the thing referred to does not change with the point of view. But in both cases we have to do with meaning differences. We shall come back to this problem in Sec. 3.3.

In mathematics and in science the assignment as well as the analysis of significance start with the rules of designation that match symbols with constructs. The second and most important step is to characterize the construct itself – and this is a matter of theory not of rule. The characterization ("definition") is often incomplete: sometimes because we want to leave room for ulterior specification, at other times because we do not know any better. The former is the case with abstract mathematics: a full specification of both sense and reference, i.e. a "definition" of the theory of a particular model, will destroy the leeway typical of abstract mathematics. In the case of science even if we wanted to effect a complete characterization we could not get it. Any characterization of the factual sense of a scientific construct is bound to be incomplete because it includes semantic assumptions that "point" to the referents without any further assistance from mathematics. When mentioning a specific domain of individuals, such as the real numbers, the mathematician can fall back on the specific theory or theories that "define" those individuals. Not so the scientist: he does not construct the referents of his theories but finds them or hopes to find them, hence he cannot fall back on other constructs.

The best he can do is to set up a mathematical formalism and conjoin it with semantic assumptions interpreting the basic concepts as so many factual items. The latter are at best described, at worst just mentioned. Whence the unavoidable meaning indeterminacy of factual theories.

2.4. *Degrees of Significance Definiteness*

There are degrees of significance definiteness corresponding to the degrees of meaning determinacy. We may distinguish the following degrees:

(i) *Low*

(*a*) Sense: definite but minimal. Reference: arbitrary except for the very general conditions set by the postulates determining the sense. Example: any abstract theory.

(*b*) Sense: not fully specified. Reference: definite. Example: any body of factual knowledge that does not contain full blown theories.

(ii) *Medium*

(*a*) Sense: minimal enriched with interpretations of some of the basic constructs. Reference: either definite or arbitrary, depending on whether or not the basic sets are specified. Example: any partial model in the sense of Ch. 6, Sec. 2.2, Definition 2.

(*b*) Sense: specified by a theory formulated in an intuitive or heuristic way. Reference: definite. Example: almost any theory in factual science.

(iii) *Large*

(*a*) Sense: nearly full. Reference: definite. Example: any theory in intuitive or nonformalized mathematics.

(*b*) Sense: specified by an axiomatic theory containing factual semantic assumptions. Reference: definite but large (genus not species of things). Example: any generic factual theory formulated axiomatically.

(iv) *Maximal*

(*a*) Sense: total. Reference: definite. Example: any fully axiomatized theory of a particular model.

(*b*) No instances in factual science.

We can now state the necessary and sufficient condition of significance. It is no more and no less than this: For a sign to be *significant* it must designate a construct (Def. 5). And for a construct to have a *reasonably definite meaning*, i.e. for being reasonably definite itself, the construct must belong to a reasonably well organized body of knowledge. Optimal meaning definiteness is attained only in a theory proper. (But maximal, i.e., full meaning is achieved only in mathematics.) While the profile of a construct can often be sketched in a satisfactory manner with more modest means, only incorporation into a theory takes automatic care of the syntax and, to some extent, of the semantics of a construct. For example mechanics will denounce 'The mass of that car is 3' as ill formed, for this expression fails to specify the mass unit. And it will expose 'Vorticity is noble' as a mongrel or semantically ill constructed expression, hence just as nonsignificant as the former expression. On the other hand ordinary language analysis has precious little to say about either except that they are grammatically sound. (Category mistakes are conceptual not linguistic: recall Ch. 2, Sec. 5.1.)

Finally a warning: We are not claiming that a construct is meaningless if it belongs to no theory, but only (i) that a construct has no *clear cut* meaning but in the midst of a theory, (ii) that a construct may *change* its meaning (i.e. it may become a different construct) if transplanted to a different theory, and (iii) that a *theoretical* construct exists only in a theory. All three aspects are illustrated in the following example. Even if different theories of social change start out with the same dictionary definition of "revolution", e.g., as "a drastic and sudden change in the established social patterns", they may elucidate or refine it in different ways, thus ending up with *different concepts* of revolution. They do this by assuming different actors (or referents of "revolution") and by emphasizing different traits as well as different causes. Thus while one theory will claim that the protagonists of revolutions (the referents of "revolution") are institutions, another will assert that they are social classes and a third that they are individuals. And while one theory will focus on institutional changes another will emphasize changes in the social and economic structure while a third view will concentrate on changes in individual roles. Finally, while one theory will assume that revolutions occur when institutions outlive their usefulness, another will claim that they are the last straw in a class struggle, and a third doctrine will assert that they happen when the members of the ruling class become corrupt. Clearly, 'revolution' signifies different constructs in the various theories of revolution even though they all share the core meaning assigned by the dictionary. (In many cases there is no such core meaning constituted by the pretheoretical or intuitive construct.) Tell me the company a construct keeps and I will tell you what it is: recall Ch. 5, Sec. 4.

3. MEANING INVARIANCE AND CHANGE

3.1. *Synonymy*

Two signs will be said to be *semantically synonymous* just in case they have the same significance. And two symbols will be regarded as *pragmatically synonymous for a given user* if he employs them interchangeably or if, under the same circumstances, they evoke in him the same reactions. Pragmatic synonymy does not presuppose semantic synonymy: thus for many people 'psychiatrist' and 'psychoanalyst' are synonymous. And few of us are consistent in matters of pragmatic synonymy. If for no other reason, semantics cannot be based upon pragmatics. A second reason is that the determination of pragmatic synonymy calls for the observation of linguistic behavior, which is irrelevant to semantic synonymy: we do not circulate questionnaires in order to find out whether 'mass' and 'inertia' are semantically synonymous. We shall restrict our considerations to semantic synonymy.

Let us begin by restating our definition in more explicit terms:

DEFINITION 7.7 Two signs are said to be *synonymous* in a given language $\mathscr{L}$ iff they have the same significance in $\mathscr{L}$:

If s and s' are in $\mathscr{L}$, then $\mathscr{S}yn(s, s', \mathscr{L}) =_{df} \mathscr{S}ig(s, \mathscr{L}) = \mathscr{S}ig(s', \mathscr{L})$.

Example 'John loves Mary' and 'Mary is loved by John' are linguistically different but synonymous sentences: they are semantically identical – they express the same proposition.

COROLLARY 7.3 Synonyms designate the same constructs:

$$\mathscr{S}yn(s, s', \mathscr{L}) \text{ iff } \mathscr{D}\text{‘}s = \mathscr{D}\text{‘}s' \text{ in } \mathscr{L}.$$

COROLLARY 7.4 Synonyms have the same vicarious sense and the same vicarious reference:

$$\mathscr{D}sc \text{ and } \mathscr{D}s'c' \text{ in } \mathscr{L} \text{ and } \mathscr{S}yn(s, s', \mathscr{L}) \text{ iff } \mathscr{S}(c) = \mathscr{S}(c') \text{ and } \mathscr{R}(c) = \mathscr{R}(c').$$

Proofs The first corollary follows from Definitions 5 and 7. The second from Corollary 3 and the axiom for ordered pairs.

Remark 1 Corollary 4 looks pedantic but it is not, for we do not know that two signs do stand for the same construct, and are thus synonymous, unless we analyze their designata into sense and reference and prove, in a more or less rigorous way, that the senses and the referents are the same. A relatively frequent case is this: two different lines of reasoning within a theory come up with a function each. Further research shows that the two functions satisfy the same differential equation and are subject to the same initial or boundary conditions. This proves that the two functions are the same, or that they differ at most by a constant. *Remark 2* Logical equivalence is insufficient for synonymy. Not even equality may be enough: thus that two functions, f and g, share their values at a point a, i.e., $f(a)=g(a)$, does not entail that '$f(a)$' and '$g(a)$' are equisignificant or synonymous. Only identity guarantees synonymy. *Remark 3* If we demand that the definiendum and the definiens have the same meaning then we can accept only identities as qualified to define, since only an identity assures us that the two sides are just different names for one and the same object. Consequences: (*a*) equivalence is not the proper form of a definition, and (*b*) the intuitive asymmetry between definiendum (*LHS*) and definiens (*RHS*) is lost in the object language: it may be regarded as a metatheoretical or as a pragmatic trait. More in Ch. 10, Sec. 2.2.

As with synonyms so with antonyms:

DEFINITION 7.8 Two signs are *antonymous* in a given language $\mathscr{L}$ iff each designates the negate of the other's designatum:

If s and s' are in $\mathscr{L}$, and $\mathscr{D}sc$ and $\mathscr{D}s'c'$, then $\mathscr{A}nt\,(s, s', \mathscr{L}) =_{df} c' = \neg c$.

Remark By the double negation theorem of ordinary logic, $\mathscr{A}nt$ proves to be a symmetrical relation. If intuitionistic logic were adopted this would be lost unless the antonymy relation were redefined. Because we find no use for intuitionistic logic in science, we won't bother with this problem.

Synonymy and antonymy are just two bands of the entire spectrum of significance relations. Both $\mathscr{S}yn$ and $\mathscr{A}nt$ come in degrees: there is weak synonymy as well as weak antonymy. Thus 'set' and 'class' are weakly synonymous, and 'war' and 'truce' weakly antonymous. The following definitions elucidate the concept of likeness of significance.

DEFINITION 7.9 If s and s' are signs designating the constructs c and c' respectively, then s and s' are *partially synonymous* (or they exhibit a *likeness of significance*) iff the intersection of the senses of c and c' is not nil: $\mathcal{S}(c) \cap \mathcal{S}(c') \neq \emptyset$.

Example 'Hemisphere' in geography does not mean the same as in anatomy, but the senses of the two constructs concerned are close even though their referents (Earth and brain) are different.

DEFINITION 7.10 If s, s' and s'' are signs designating the constructs c', c' and c'' respectively, then the *likeness in significance* between s and s' is *closer* than the likeness in significance between s and s'' iff $\mathcal{S}(c) \cap \mathcal{S}(c') \supset \mathcal{S}(c) \cap \mathcal{S}(c'')$.

The preceding elucidations of the notions of equality and likeness of meaning should meet Quine's objection to propositions, namely that "If there were propositions, they would induce a certain relation of synonymy or equivalence between sentences themselves: those sentences would be equivalent that expressed the same propositions" (Quine, 1970b, p. 3, also 1960, Ch. VI). Well, don't they?

Our treatment of meaning relations, while at variance with the lexical view of semantics (Katz and Fodor, 1963), is in agreement with the view (Bar-Hillel, 1970) that such relations are logical relations. And our definition of synonymy confirms the view that the identity of the sets of implicants and implicates (i.e. sense) is necessary but insufficient for synonymy (Attfield and Durrant, 1973).

3.2. *Meaning Invariance*

The relation $\mathcal{S}yn$ of synonymy introduced by Def. 7 (Sec. 3.1) is an equivalence relation. Hence it defines the equivalence classes constituted by synonymous signs. In other words, for any language $\mathcal{L}$, if s is in $\mathcal{L}$,

$$[s]_{\mathcal{L}} = \{t \in \mathcal{L} \mid \mathcal{S}yn\,(s, t, \mathcal{L})\}$$

is the set of all the synonyms of s in $\mathcal{L}$. Every one of these equivalence classes corresponds to a single construct.

Take now the totality of equivalence classes under $\mathcal{S}yn$, i.e. the quotient set $\Sigma^{**}/\mathcal{S}yn$. This is the linguistic representative of all the constructs expressible in $\mathcal{L}$. Even if the designation function $\mathcal{D}$ is many-one, as we have assumed, we now have a one-one function $\mathcal{D}^*$ that matches distinct elements of $\Sigma^{**}/\mathscr{S}yn$ to distinct members of the set C of constructs expressible by $\mathscr{L}$. This isomorphism $\mathscr{D}^*: \Sigma^{**}/\mathscr{S}yn \to C$ may be called the *regular designation function*.

Let us pull the various items together. We start with the many-one function (surjection) $\mathscr{D}: \Sigma^{**} \to C$. Then we define the equivalence relation $\mathscr{S}yn$ in Σ^{**}. This relation determines the projection $p: \Sigma^{**} \to \Sigma^{**}/\mathscr{S}yn$ that assigns to every symbol the class of its semantic equivalents. Next we map the range of p onto C. Finally we compose p with $\mathscr{D}^*$. The outcome is $\mathscr{D}^* \circ p = \mathscr{D}$ as pictured by the diagram

$$\begin{tikzcd}
\Sigma^{**} \arrow[rr, "\mathscr{D}"] \arrow[dr, "p"'] & & C \\
& \Sigma^{**}/\mathscr{S}yn \arrow[ur, "\mathscr{D}^*"'] &
\end{tikzcd}$$

Now pick all the sentences of $\mathscr{L}$. Next group them into classes of synonyms, i.e., equisignificant sentences. Finally allow $\mathscr{L}$ to range over the set of all possible languages. That is, build the family of equivalence classes of sentences under the relation of equisignificance:

$\{[s]_{\mathscr{L}} \mid s \text{ is a sentence of } \mathscr{L} \;\&\; \mathscr{L} \text{ is a conceptual language}\}$. A nominalist might wish to identify this family of equisignificant sentences with what we call a *proposition* (or statement). But he cannot do this because he has no independent criterion of significance. In other words, the above cannot be treated as *defining* "proposition" because the latter concept is involved in the formation of the equivalence classes $[s]_{\mathscr{L}}$. Indeed, as we saw in Sec. 3.1, we do not know whether two sentences in mathematics or in science are synonymous unless we can show that they designate the same construct. Still, the above does clarify the idea that a proposition is that which remains invariant under all faithful sentence to sentence translations, as suggested by Russell (1940).

The translation concept can be elucidated in the following way. Consider the set $S_{\mathscr{L}}$ of all the sentences of a language $\mathscr{L}$ and the homologous collection $S_{\mathscr{L}'}$ for another language $\mathscr{L}'$. Even if these two sets are disjoint and structurally different, there may exist a one many relation

τ from $S_{\mathscr{L}}$ to $S_{\mathscr{L}'}$ that preserves significance. If such a relation exists we say that τ is a translation of $\mathscr{L}$ into $\mathscr{L}'$. More explicitly, we have

DEFINITION 7.11 Let $\mathscr{L}$ and $\mathscr{L}'$ be two conceptual languages and let τ be a one many relation from the set of sentences $S_{\mathscr{L}}$ to the set of sentences $S_{\mathscr{L}'}$. Then τ is said to be an *exact pointwise translation* of $\mathscr{L}$ into $\mathscr{L}'$ iff

$$\mathscr{S}\!\mathit{ig}(\tau\text{‘}s) = \mathscr{S}\!\mathit{ig}(s) \quad \text{for every} \quad s \in S_{\mathscr{L}},$$

where $\tau\text{‘}s \in S_{\mathscr{L}'}$ is a translation of s into $\mathscr{L}'$.

This concept of translation is useful in mathematics, where it can be strengthened to that of a function (Wang, 1951). But it is hardly applicable to natural languages, where not all sentences are independently meaningful. Here whole clusters of sentences must be paired off if significance is to be preserved. In other words, in the case of natural languages we must give up the ideal of pointwise translation and settle instead for global translation. However, this need not force us to adopt Quine's doctrine of the inevitable indeterminacy of translation: it only suggests supplementing Definition 11 with the following elucidation of the notion of global translation:

DEFINITION 7.12 Let $S_{\mathscr{L}}$ and $S_{\mathscr{L}'}$ be the sets of all sentences of languages $\mathscr{L}$ and $\mathscr{L}'$ respectively, and call $\mathscr{P}(S_{\mathscr{L}})$ and $\mathscr{P}(S_{\mathscr{L}'})$ their corresponding power sets. Then τ is an *exact global translation* of $\mathscr{L}$ into $\mathscr{L}'$ iff τ is a relation from $\mathscr{P}(S_{\mathscr{L}})$ to $\mathscr{P}(S_{\mathscr{L}'})$ such that

$$\mathscr{S}\!\mathit{ig}(\tau\text{‘}u) = \mathscr{S}\!\mathit{ig}(u) \quad \text{for every} \quad u \in \mathscr{P}(S_{\mathscr{L}})$$

where $\tau\text{‘}u \in \mathscr{P}(S_{\mathscr{L}'})$ is a translation of the set of u of sentences of $S_{\mathscr{L}}$ into $\mathscr{L}'$.

In sum, our concept of synonymy has allowed us to define two concepts of translation: pointwise and global. These concepts are strictly semantic as it should be: translation is concerned with meaning not with structure. (For the opposite view and a purely syntactic definition of translation see Svenonius (1973).) Surely perfect translations, even if global, are hard to come by. But, since they are desirable, the above definitions, far from being idle, may perform a regulative function. Thus it would be desirable to regularize the grammars and vocabularies of some of the natural languages so as to render them perfectly translatable into one another. Once this linguistic reform were implemented machine translation should offer no obstacles.

The concept of translation is relevant to linguistics, the foundations of mathematics, and the foundations of science – where it occurs with reference to equivalent theories utilizing different mathematical "languages". (For the use of theories as languages for further theories, see Ch. 1, Sec. 2.3.) However in factual science one is much more interested in different theories, whether or not they are couched in the same mathematical "languages". And, unlike translation, the passage from one theory to another may involve meaning changes. This matter deserves a separate subsection.

3.3. *Meaning Change*

If meaning is *sensus cum referens* then a change in meaning is a change in either sense or reference or both. And either change, when stripped of its pragmatic aspects, consists of a *difference* in sense and/or reference. Since both sense and reference have been constructed as sets, it is natural to define the difference in either as a symmetric (or Boolean) difference. It is equally natural to define difference in meaning as the pair ⟨difference in sense, difference in reference⟩. More precisely we have

DEFINITION 7.13 Let c and c' be constructs. Then

(i) The *difference in sense* between c and c' is $\delta_{\mathcal{S}}(c, c') = \mathcal{S}(c)\,\Delta\mathcal{S}(c')$;

(ii) the *difference in reference* between c and c' is $\delta_{\mathcal{R}}(c, c') = \mathcal{R}(c)\,\Delta\mathcal{R}(c')$;

(iii) the *difference in meaning* between c and c' is $\delta_{\mathcal{M}}(c, c') = \langle\delta_{\mathcal{S}}(c, c'), \delta_{\mathcal{R}}(c, c')\rangle$.

The least interesting case is the one in which the meaning "distance" between two constructs is maximal:

COROLLARY 7.5 Let c and c' be semantically unrelated constructs, i.e. such that both $\mathcal{S}(c) \cap \mathcal{S}(c') = \emptyset$ and $\mathcal{R}(c) \cap \mathcal{R}(c') = \emptyset$, which amounts to $\mathcal{M}(c) \times \mathcal{M}(c') = \langle\emptyset, \emptyset\rangle = \Box$. Then

$$\delta_{\mathcal{M}}(c, c') = \langle\mathcal{S}(c) \cup \mathcal{S}(c'), \mathcal{R}(c) \cup \mathcal{R}(c')\rangle.$$

The next case is far more interesting, as it concerns constructs with comparable senses:

COROLLARY 7.6 Let c and c' be two constructs such that the sense of c' contains the sense of c and such that they are coreferential. In brief, assume $\mathcal{S}(c')=\mathcal{S}(c)\cup\Delta\mathcal{S}$, with $\Delta\mathcal{S}\neq\emptyset$ and $\mathcal{S}(c)\cap\Delta\mathcal{S}=\emptyset$, and $\mathcal{R}(c)\cap\cap\mathcal{R}(c')\neq\emptyset$. Then

$$\delta_{\mathcal{M}}(c, c')=\langle\Delta\mathcal{S}, \Delta\mathcal{R}\rangle.$$

Example 1

$c=$ Partially ordered set $\langle A, \preccurlyeq\rangle$
$c'=$ Semilattice $\langle A', \preccurlyeq, \wedge\rangle$
$\mathcal{S}(c)=$ Axioms and theorems for posets, $\mathcal{R}(c)=A$
$\mathcal{S}(c')=\mathcal{S}(c)\cup$ Axioms and theorems containing $\wedge$, $\mathcal{R}(c')=A'\subseteq A$
$\delta_{\mathcal{M}}(c, c')=\langle$Axioms and theorems containing $\wedge$, $A-A'\rangle$.

Example 2

$c=$ Celestial mechanics (CM)
$c'=$ Lunar theory (L)
$\mathcal{S}(c)=$ Axioms and theorems of CM, $\mathcal{R}(c)=$ All celestial bodies
$\mathcal{S}(c')=\mathcal{S}(c)\cup$ Hypotheses concerning the Moon only, $\mathcal{R}(c')=\{$Moon$\}$
$\delta_{\mathcal{M}}(c, c')=\langle$Hypotheses concerning the Moon only, All celestial bodies except the Moon$\rangle$.

Whether or not our concept of meaning change elucidates the intuitive ideas put forth by Hanson, Kuhn, Toulmin and Feyerabend, and tossed so vehemently by so many philosophers, is hard to say. The preceding elucidations are offered as a semantic framework within which historical examples can be profitably discussed. Of course one may get inspiration for a theory of meaning by studying case histories: but such a study does not constitute an analysis of meaning, let alone a theory of meaning. Without a previous agreement on what 'meaning' signifies, i.e., unless a definite theory of meaning (and meaning change) is shared, if only for the sake of argument, the latter is bound to be chaotic, hence sterile. (For an example of such a dialogue among the deaf see the Minnesota discussion on correspondence rules, in Radner and Winokur (1971). For criteria for meaning changes see Kleiner (1971).)

The study of actual changes in significance belongs to pragmatics, historical linguistics, and the history of ideas. From this viewpoint every sign has a certain flexibility, also called *open texture* (Waismann, 1955). Thus "solid" has been redefined a number of times and will presumably be the object of further elucidations as the theories of solids evolve. It is only within the symbolism of a given theory that signs are nearly nonporous. Nevertheless the various significances assigned a scientific term in the course of its history often do have a solid core – namely the intersection of all its various significances. This core is not the "essence" of the sign: it may well consist of certain external characteristics. And it is often minute.

4. FACTUAL AND EMPIRICAL MEANINGS

4.1. *Definitions*

To elucidate the notion of factual meaning we blend the results of Ch. 2, Sec. 4.1 concerning factual reference with those of Ch. 5, Sec. 3.3 about factual sense. We thus get a specialization of Def. 1 in Sec. 2.1:

DEFINITION 7.14 Let c be a construct with a factual sense $\mathcal{S}_F(c)$ and a factual reference $\mathcal{R}_F(c)$. Then the *factual meaning* of c is defined as

$$\mathcal{M}_F(c) = \langle \mathcal{S}_F(c), \mathcal{R}_F(c) \rangle.$$

Example 1 c = Electrodynamics, or e for short.

$$\mathcal{M}_F(e) = \langle \{\text{Law statements, meaning assumptions, etc., of } e\}, \text{Electromagnetic fields} \cup \text{Bodies} \rangle.$$

Example 2 c = Mind concept, or m for short.

$$\mathcal{M}_F(m) = \langle \ulcorner \text{The internal activity of the brain} \urcorner, \text{Higher animals} \rangle.$$

Factual meaning should not be mistaken for empirical meaning. A construct may be said to have an empirical meaning just in case it refers at least partially to human experiences of some kind – e.g. perceiving, thinking, or doing. Thus while ⌜There are neutrinos⌝ is factually meaningful (and even true), it is empirically meaningless, for we have no experience of neutrinos. If a construct is empirically meaningful then it is factually meaningful – but not conversely. This principle of our semantics matches the metaphysical assumption that experience is a proper part of reality – namely the one involving experients. Both principles are of course at the center of a realist philosophy, of which more in Ch. 10, Sec. 3.3.

The difference between the factual and the empirical can be rendered slightly more precise by introducing the following conventions.

DEFINITION 7.15 A predicate $P: A \times B \times \ldots \times N \to S$, where S is a set of statements, is called *factual* iff at least one of the cartesian factors of the set on which P is defined represents a domain of factual items.

DEFINITION 7.16 A factual predicate $P: A \times B \times \ldots \times N \to S$ is called *empirical* iff at least one of the cartesian factors of the set on which P is defined is a set of sentient organisms.

DEFINITION 7.17 A predicate that is factual but not empirical is called *strictly factual* or *objective*.

Example Whereas "temperature" is strictly factual or objective, "hot" is empirical for having been defined on the set of ordered pairs thing-sentient being.

The rest is obvious. A statement is factual iff it contains at least one factual predicate, empirical iff it contains at least one empirical predicate, and strictly factual iff it contains factual but not empirical predicates. Similarly for sets of statements, in particular for theories.

Warning: Scientists sometimes call 'meaningless' what is actually meaningful but either uninteresting or false. For example, the runaway solutions of the equations of motion of classical electrodynamics are sometimes said to be 'physically meaningless'. As a matter of fact they are meaningful: they represent the motion of a self-accelerated point charge. Only, they are false. Moral: Dig out the concepts underneath the words.

4.2. *The Search for Factual Meaning*

When a scientific theory attains axiomatic maturity, the basic constructs determine the meanings of all others. This state of refinement is of course an outcome of a process of concept and theory formation that is anything but rule directed. In the preaxiomatic stages, i.e., in all theoretical domains save foundations research, the search for meaning is, like the search for hypotheses and theories, a guessing-trial-correction zigzagging. Even if the mathematical ideas are clear, their factual sense and occasionally even their referents are all too often imprecise at this stage. In sum, the semantics of a factual theory, i.e., its sense and reference, emerges gradually. It grows as a result of (i) solving more and more problems in the theory, (ii) enhancing the organization of the theory, (iii) relating the theory to other theories, and (iv) analyzing and evaluating the key constructs of the theory.

A typical situation in the search for the factual meaning of a theoretical construct is this:

(i) formulating a problem in the context of a given theory (very often an ill organized one);

(ii) detaching the mathematical component of the problem – i.e. formulating a problem in mathematics;

(iii) solving the mathematical problem;

(iv) wondering about the factual meaning of the solution.

This last task can be very hard, particularly in the preaxiomatic stages. Even if all the constructs occurring in the formulation of the problem have clear cut meanings, the solution may be quite unperspicuous: we may be able to "read" every symbol in it without however "making sense" out of the whole. For what we really are after is not a term by term interpretation: what we want to find out is what the solution *represents* – i.e. what aspects (e.g. properties) of the system it symbolizes and what facts (e.g. events), if any, it models. This is why J. C. Maxwell claimed that v^2, the square of a particle velocity, "has no distinct physical meaning", nor does mv^2, where m represents the particle mass (Maxwell, 1871). Actually the compound construct "mv^2" is perfectly meaningful in our large sense of 'meaning', since it is built out of individually meaningful constructs in a formally correct way. The trouble with it is not that it is meaningless but that it fails to *represent a definite property* of the system concerned. On the other hand v and mv do represent a property each and so does $\frac{1}{2}mv^2$. Why should the factor $\frac{1}{2}$ make such a huge semantic difference? Because it is $\frac{1}{2}mv^2$, not mv^2, that figures as an independent term (an addend) in a *law* statement of particle mechanics, i.e. in a theorem that is supposed to represent a natural pattern. In short, although both "mv^2" and "$\frac{1}{2}mv^2$" have definite senses and the same referent, the former is attributed no "distinct physical meaning" in the sense that it represents no particular property of the system concerned – and this in turn because it plays no role in any law statement.

The occurrence of a construct as an identifiable component in a law statement (e.g., as an addend) is then a good clue as to its factual meaning. No less and no more. Indeed, one and the same "quantity" (magnitude) can often be decomposed in different ways and these differences are only mathematically meaningful. Moreover, even if a construct occurs as a separate component in a law statement, we may fail to "read" it without further ado and so may have to resort to other procedures. For example, in a field theory a construct such as $\nabla \cdot V$, where V is a vector field, may not be "identified" or "recognized", i.e. assigned a "distinct physical meaning" right away. We may first have to integrate $\nabla \cdot V$ over a space region: by Gauss' theorem the outcome will be the flux of V across the boundary of that region, and this derived quantity may represent a property of the system. But even this may be insufficient: we may have to look for additional clues. A very fruitful one is dimensional analysis. Thus if the dimension of a magnitude X is LT^{-1} we may suspect that X represents the velocity of something. But then it may not.

Very often a theoretical construct with a perfectly definite referent has no manifest factual meaning – or, if preferred, its meaning is hidden. This is particularly the case with what may be called the *source magnitudes*. These are functions with definite reference classes and definite senses but representing no definite property of their referents although they generate a number of representing functions. Examples: (*a*) the various potential functions, whose gradients represent forces; (*b*) the partition function in statistical mechanics which, by mathematical manipulation, yields a number of representing functions; (*c*) the wave function and the statistical operator in the quantum theories. As long as a source function furnishes functions representing one property each we should tolerate, nay encourage them and defend them from the attacks of the operationist, who has no use for them because they are not directly measurable.

Gradually and somehow or other, both the mathematical formalism and its factual meaning mature to the point that the theory becomes ripe for axiomatization. In particular, the semantic assumptions can be stated explicitly and so contribute, alongside the remaining assumptions, to delineating the factual meaning of the theory. Once the theory has been given an axiomatic formulation everything should run smoother than before. For in a well organized theory everything flows from the axiom top: both theorems and meanings. In principle the meaning of a defined construct will be found by analyzing its definition and the meaning of a theorem will be found by analyzing its component constructs as well as the premises that entail it. In principle but not necessarily in practice: axiomatization eases proofs and interpretations and makes them more precise but not mechanical.

4.3. *Shape and Role of Meaning Assumptions*

There is no consensus on how to handle the factual meaning of scientific constructs. Every scientist proceeds in his own way, though often influenced by some philosophical school. Here follow in quick succession the main current views on the matter – or rather sketches of them.

(i) *Formalism and conventionalism.* No meaning assumptions are called for because scientific constructs have no factual meaning: they are just pieces in a mathematical machinery. A scientific theory is identical to its mathematical formalism. *Criticism* Any mathematical formalism may be assigned alternative interpretations: a factual construct is a mathematical construct together with a factual interpretation (Ch. 6, Sec. 3).

(ii) *The semantic miracle view.* Meanings take care of themselves: every formalism generates its own interpretation. Hence no meaning assumptions are needed. *Criticism* Same as for (i). The reason some formalisms seem to be necessarily associated with certain interpretations is just habit. A scientific specialist is so used to handling certain form-content pairs that it may not occur to him that the same form may be paired to an altogether different content.

(iii) *The meaning-lies-in-the-name view.* All we need to turn a mathematical formalism into a scientific theory is to add designation rules, such as ⌜The parameter t is called time⌝. *Criticism* While designation rules are necessary ingredients of the semantics of a scientific theory, they are insufficient. Names are conventional, meaning assumptions are not: the latter are testable, hence capable of being overthrown. Thus although we still use the electric current concept we no longer assume that it represents the rate of flow of the electric fluid. If we claim that a theory represents certain things then we must state clearly what in the theory represents what in the world.

(iv) *Classical empiricism.* Meanings are assigned by ostensive definitions or rules, such as "That is blue". *Criticism* First, an ostensive rule may be a subject for pragmatics but not for semantics: unless accompanied by suitable gestures it is devoid of (pragmatic) meaning. Second, unfortunately for toddlers, the most interesting constructs in science and even in ordinary knowledge are non-ostensive.

(v) *Operationism.* Meanings are assigned by specifying modes of observation, measurement, or action in general. For example, the concept of a thermodynamic state must be specified by describing a method of preparation of states (Carathéodory, 1924; Giles, 1964). *Criticism* First, most of the referents of a scientific theory are beyond the reach of the experimentalist – if only because they are possibles rather than actuals. Hence if a thermodynamic state must be the outcome of a human act then most physical systems are in no state at all. Second, this view is dangerously near subjectivism, which has no need for empirical procedures. Indeed, once a thermodynamic state has been made to depend on man, why not on mind alone? This step has actually been taken: it has been claimed that "a state is a state of mind induced by the available knowledge of preparation" (Burton, 1968). Third, the whole of operationism rests on a confusion between reference and evidence (Feigl, 1958; Bunge, 1967a, 1973a, 1973b).

(vi) *Mellowed operationism.* A theoretical term is significant to the extent to which it is related, in the body of a theory, to some observational sentence of the theory (Carnap, 1956). More precisely, "A theoretical term t is significant if there is an assumption A involving t such that from A and additional assumptions involving other theoretical terms that have already been recognized as significant it is possible to derive with the help of the postulates and the rules of correspondence an observation sentence that cannot be derived without the assumption *A*" (Carnap, 1963a, p. 80). *Criticism* First, most scientific theories contain no observational constructs. Strictly speaking *all* the constructs in a scientific theory are theoretical and often devoid of empirical (yet not of factual) meaning. Hence the theoretical/observational dichotomy is not applicable to scientific theories. Second, as suggested by Carnap himself, testable statements can only be had with external help – namely by roping in assumptions belonging to other theories as well as empirical evidence. (For details see Bunge (1967a) and (1970a).) Even so, this will render the theory empirically testable but not empirically meaningful.

(vii) *Realism.* Factual theories have a factual meaning. And factual meanings are determined jointly by all the assumptions of the theory – in particular, though not exclusively, by the semantic assumptions. Since the latter bear on the basic concepts, they belong in the axiomatic foundations of the theory. Every semantic assumption indicates the referent(s) of the construct and suggests what it represents. For example, a conditional probability value $P_n(b \mid a)$ can be made to represent the tendency of the *n*th system of a certain kind to jump from state *a* to state *b*. It makes no difference to the semantics of a theory whether the referents are perceptible and the traits that are being represented by the theory are directly observable or must be reached with the help of further theories: observability is relevant to testability not to meaning. In other words, the semantic assumptions in a scientific theory relate constructs to things and some of their traits: anything that relates constructs to empirical operations, such as preparation or measurement, may qualify as a *testability condition* but certainly not as a meaning assumption. *Criticism* None that I can think of.

Every one of these policies has its advocates and all of them but classical empiricism (ostensivism) are practised nowadays by scientists. Yet popularity is not the seal of truth: just because most scientists either do not care to formulate any explicit meaning assumptions or because they propose operationist interpretations, does not prove that those policies are right. A semantic policy, just as any other policy, must be judged by its success in achieving the goals it sets itself as well as by its success in surviving criticism. If judged in this dual way, the most popular semantic policies prove to be utter failures. The first three because they do not even attempt to pinpoint the peculiarities of factual theories *vis-à-vis* their mathematical formalisms. And the two varieties of operationism are flops as well because the semantic assumptions they recommend are too narrow: they are linked to particular laboratory techniques, hence closed to possible alternatives. And most of the time they are phony, as they stipulate impossible measurements. Try measuring a lagrangian density, or a partition function, or a state function.

The failure of the various attempts to specify the meanings of theoretical concepts by reducing, or at least linking them to observational concepts, has produced a mood of despondency among philosophers of science. Thus Putnam came to the conclusion that the very *problem* of interpreting theoretical terms "does not exist" (Putnam, 1962) and Hempel thinks now that the problem was "misconceived" (Hempel, 1970). Nevertheless the problem won't lie down: the theoretical scientist faces it every day when pondering over the possible factual meanings of his mathematical formulas. Moreover the scientist tries to solve his semantic problems in more of less naive ways, without the benefit of a full fledged semantic theory. No doubt he would fare better if, instead of being told that he should worry no more because he has really no problem at all, he were offered a definite semantic theory.

The failure of empiricist semantics does not entail the impossibility of every kind of semantics. It just suggests looking elsewhere for a solution to the genuine and difficult problem of specifying (or rather sketching) the meaning of theoretical concepts (or the significance of theoretical terms). An alternative to empiricism is realism, the only semantic policy that has emerged unscathed from the fifty years' war over the soul of scientific constructs. True, realism is not often practised in an explicit manner, i.e., by laying down the rules of denotation and the semantic assumptions that delineate the meanings of the undefined concepts of a scientific theory. But then (*a*) impopularity is not the seal of falsity, (*b*) few theoreticians care to render all their assumptions explicit, and (*c*) positivist semantics, even after having been repudiated by its propounders, still enjoys a considerable prestige among scientists. It behooves the philosopher to illumine the path by showing in particular cases how to interpret theoretical concepts in terms of facts.

We close with a warning. A semantic assumption occurring in an axiom system could rightly be called a *meaning postulate*. Unfortunately this expression was preempted by Carnap (1952), who employed it in a different sense. Consider the two standard examples of a "meaning postulate", first of all

⌜For every *x*, *x* is a bachelor iff *x* is a male and *x* is not married.⌝ (1)

This is a construct-construct relation that does not intend to say what "bachelor" refers to and represents but rather how it is related in extension to both "male" and "married". It seems to be an ordinary dictionary definition. In any case it is supposed to be a priori, hence empirically unassailable. The second standard example of a "meaning postulate" is

⌜For every *x*, if *x* is a bachelor then *x* is not married.⌝ (2)

This might be taken as a "law" of ordinary knowledge or as a linguistic fact or, finally, as a deductive consequence of the convention (1). In neither case is it an empirically refutable semantic assumption or axiom in our sense. Nor is it a postulate in the sense of 'axiom' (Carnap, 1952). Whatever it is, a Carnapian "meaning postulate" does not contribute to the sense of a scientific term, hence it takes no part in the axiomatic reconstruction of a scientific theory, whence it need not concern us.

5. MEANING ET ALIA

5.1. *Meaning and Testability*

We have elaborated the common sense or realist view of factual meaningfulness. According to this view a statement is factually meaningful only if it refers to, and the more so if it represents, a factual item – a thing, a state of affairs, or an event. The referent need not be actual and the representation, if there is any, need not be true: the statement may well concern the past or the future, and it may be utterly false or even impossible to test. Any narrower criterion of factual meaningfulness runs the risk of disowning the most interesting scientific speculations.

This commonsensical view was also the one Carnap held before he came under the influence of Wittgenstein: "The meaning of a statement lies in the fact that it expresses a (conceivable, not necessarily existing) state of affairs" (Carnap (1928) in Carnap, 1967, p. 325). Eventually

Carnap and the other members of the Vienna Circle succumbed to Wittgenstein's verifiability doctrine of meaning, according to which the meaning of a sentence consists in the way it can be verified – i.e. in its truth conditions. (Cf. Schlick, 1932/33.) If a sentence belongs to empirical science, the truth conditions must describe procedures of empirical test. Whence *Meaning = Testability*. And because only observational statements are empirically testable, *Meaning = Observability*. Put negatively: Whatever sentence is not verifiable by observation is empirically (= cognitively) meaningless.

This strict thesis was somewhat liberalized later on but without relinquishing the dependence of meaning on testability. The criterion that ultimately prevailed in the logical empiricist camp was this: A sentence is empirically (or cognitively) significant only if its sole extralogical constants are observational (e.g. 'sticky' and 'smelly') or if, conjoined with further sentences, it entails observational sentences (Carnap, 1956, 1963a; Rozeboom, 1962). In the latest version of this thesis testability need not be scientific hence objective: "I regard as meaningful for me whatever I can, in principle, confirm subjectively" (Carnap, 1963b, p. 882). Once again the subjectivist component of empiricism won out.

We criticized earlier the identification of meaning with testability (Ch. 4, Sec. 3.3). Suffice here to remark that false statements are as meaningful as true ones, and that all scientific theories contain incompletely testable or even wholly untestable statements: for example, quantum mechanics allows one to compute the velocity of an electron in an atom, which is an empirically inaccessible function. Let us undertake to state, even if in summary, the actual relations between meaning and testability.

It is a rule of scientific method to abstain from assigning truth values, except for the sake of argument, until after the relevant empirical evidence is at hand. In other words, testing is necessary for truth value: *Truth value ⇒ Testing*. In turn, a necessary condition for any actual tests is that the statement concerned be testable, i.e. that the theories and the empirical techniques of the day judge the statement to be susceptible of being confronted with facts, if not right now later on. In sum, *Testing ⇒ Testability*. Now, if a statement is susceptible to test then it is meaningful to begin with, i.e. it has a nonempty sense and a reference class – which may, alas, prove to be empty. Otherwise it would be impossible to devise a test for the statement. Briefly, meaningfulness, though insufficient for testability, is necessary for it: *Testability* ⇒ *Meaningfulness*. Finally if a statement is factually meaningful, i.e., if it has a factual sense and a factual reference (whether actual or possible), then it is well formed, i.e. syntactically meaningful in some formalism. (For a definition of the latter notion of meaningfulness see Tarski (1956), p. 284.) In short, *Meaningfulness* ⇒ *Well-formedness*.

In conclusion, the complete *logical* chain is

> *Truth value* ⇒ *Testing* ⇒ *Testability* ⇒ *Meaningfulness* ⇒ *Well-formedness*.

Consequently the *methodological* sequence is

> *Checking well-formedness* → *Assigning or analyzing meaning* → → *Judging testability and devising tests* → *Actual testing* → *Assigning truth value.*

5.2. *Meaning and Use*

At about the same time as Carnap was working out the *Tractatus*' view on meaning, Wittgenstein was busy demolishing it and attempting to sketch a pragmatist philosophy of language. On this view language is only a social activity and the meaning of an expression consists in its use. In turn, uses are established by custom as recorded by the (Oxford) dictionary – not by any theoretical analysis. As one of the disciples put it, "To give the meaning of an expression (in the sense in which I am using the word) is to give *general directions* for its use in making true or false assertions" (Strawson, 1950). Therefore "Ultimately a meaning-statement (a statement as to what a linguistic expression means) is to be tested by determining what people do in their employment of the expression in question" (Alston, 1968, p. 145).

What the second Wittgenstein and his apostles were interested in was, of course, the notion of *pragmatic meaning*. This concept can be exemplified but has so far escaped theoretical elucidation. (Only the notion of pragmatic synonymy has been elucidated – not by the Wittgensteinians but by Carnap (1939) and Naess (1956).) In any case it is different from the concept of semantic meaning and no possible substitute for the latter. The same "general directions" for manipulating a bunch of symbols are consistent with alternative significations assigned to those symbols. For this reason linguistic pragmatism cannot explain why two quantum physicists, while disagreeing on the significance of the symbols they use, can arrive at the same formulas. While linguistic pragmatism is incompetent to tackle semantic meaning, it can be effective in misleading philosophers into thinking that meanings can only be discovered by listening to (ordinary) talk rather than by uncovering sense and reference. Surely field observations can disclose pragmatic meanings – provided they are conducted with the methodological apparatus of the linguist. It is linguists not philosophers who are equipped to carry out linguistic investigations. Philosophers should philosophize about linguistics among other things – not language.

5.3. *Meaning and Understanding*

The significance of a sign and the meaning of the construct it designates must not be mistaken for the mental process of understanding either. Meanings are supposed to be objective whereas the experience of thinking of them is, like any other experience, subjective. This distinction, which goes against the grain of empiricism, goes back to Bolzano, Lotze, Frege, and Meinong. It is consecrated by the distinction, emphasized in the Introduction to Volume 1, between philosophical semantics and the psychology of cognition and language, which is a branch of factual science.

The distinction between meaning and the understanding of meaning is correct provided meanings are not reified or turned into Platonic Ideas. With meanings as with their carriers, i.e., constructs, we may adopt a fictionalist stand: we may *pretend* that they exist – without however assuming that they have an *autonomous* existence. No rational beings, no constructs; no constructs, no meanings. We have no use for the idealist thesis that there are concepts and propositions in themselves, i.e. apart from thinking beings. But if we wish to build a *theory* of constructs that brushes aside the particular ways and circumstances in which constructs are being thought, then obviously we must abstract from such ways and circumstances. Such a theory cannot possibly enter into conflict with any psychological theory concerning the understanding (or the misunderstanding) of a meaning, since it will not pose the question.

Although the concepts of meaning and of understanding belong to different fields of research, they are related, namely thus: *If anything* *is understandable then maybe it is meaningful.* ('Maybe' because any claim to intelligibility is of uncertain value.) Practical consequence: The clearer and more orderly the presentation of a body of knowledge the better the chances of understanding it. It does not follow that axiomatics is the ideal didactical wrapping. It does follow that a judicious combination of axiomatics with intuitive remarks and a statement of motivations is the best one can do to *facilitate correct understanding.* (See Bunge, 1973b, Ch. 8, Sec. 6.)

Not surprisingly, recent work in psycholinguistics and in artificial intelligence bears out our thesis that isolated sentences are nonsignificant, hence unintelligible. In fact to understand a sentence a person (or a computer) must know the language the sentence belongs to, must be able to do some reasoning, and must have some substantive information (Winograd, 1972).

A last question: Can computers grasp any meanings? The short answer is: No, because they handle physical signals not constructs. It is the programmer who assigns to these signals definite constructs, hence the meanings of the latter: he does this when setting up or using his programming code. In particular when he reads (interprets) the computer output – or, for that matter, when he reads anything. Unlike its programmer, the computer does not have to and cannot possibly interpret anything. So much so that a computer is incapable of making interpretation errors: only rational beings can make semantic mistakes. Computers are used not because they replace minds but because they simulate some aspects of the human mind. Only a living brain can have a whole mind of its own. And only some brains pose new conceptual problems, invent theories, and evaluate the latter. For a computer to understand a sign (e.g. a sentence) consists in associating it to the correct state (of the machine) to produce a final output (or machine state) according to a definite rule included in the program. No understanding of meaning is involved in this purely physical operation.

5.4. *Factual Meaning and Covariance*

Place words like 'here' and time words like 'now' do not have the same significance for everyone: they are subject-dependent or egocentric. Similarly space and time coordinate values are local not universal. Hence a statement such as ⌜Particle p is at place x at time t⌝, even if true relative to some reference frame, is not universally true, i.e., true in, or relative to, every possible reference frame. (But if we know how the given frame is related to some other frame then we can translate the statement into another statement that will hold in, or relative to, the new reference frame. The Galilei and the Lorentz transformation formulas are two such translation devices. Recall Ch. 3, Sec. 2.3).

In other words, position and time values are not *invariant* under every coordinate transformation. (On the other hand electric charge and transition probability values are invariant.) And statements involving position and time coordinates are not always *covariant* (or form invariant) under certain coordinate transformations. In either case, what is at stake is the permanence, or lack of it, under changes in the mode of representing or mapping facts in spacetime: invariance (or noninvariance) in the case of properties, covariance (or absence of it) in the case of law statements. For example, a Galilei covariant formula holds in every Galilei frame, i.e. it is covariant under the group of Galilei transformations. This is the case with Newton's basic equations of motion, not however with their solutions: the latter are frame-dependent – but at least we know how to translate them into statements that are true in, or relative to, alternative frames. But Newton's equations are not covariant under different coordinate transformations such as Lorentz'.

It is tempting to conclude that only invariant properties and covariant equations have a factual or objective meaning, all others being endowed with a subjective meaning if any. Thus Weyl (1919, p. 129) held that a relation between points in spacetime has an objective meaning (*objektive Bedeutung*) just in case it is invariant under Galilei's transformations. But then Galilei's law of falling bodies, a special solution of Newton's equations of motion, would be devoid of objective meaning because it is frame-dependent (not covariant). This is an abuse of the word 'meaning'. What is really at stake is *truth* not meaning: A statement involving spatial or temporal concepts (e.g., spacetime coordinates) and making explicit reference to some frame can be fully meaningful without being *universally true*, i.e., true in (relative to) every possible frame.

But what is invariant (or else covariant) under a certain transformation group may cease to be so under a different group. Thus while the relation of simultaneity is Galilei invariant, hence absolute in Newtonian physics, it is assumed to be frame-dependent, hence relative, in relativistic physics. In the latter context, therefore, a statement of the form ⌜Events *a* and *b* are simultaneous⌝ is regarded as being ill formed unless the context makes it clear which reference frame has been adopted. Relative to a certain frame f it may be that the two events are simultaneous, but then they may fail to be simultaneous relative to an alternative frame f'. Consequently the corresponding statements will not be universally true but instead locally so. In obvious symbols: while $\ulcorner S(a, b, f)\urcorner$ may be true, hence meaningful, $\ulcorner S(a, b, f')\urcorner$ may be false, hence meaningful as well. On the other hand Weyl inferred that simultaneity has no objective meaning (Weyl, *op. cit.*, p. 146). Which, in addition to being a semantic mistake, contradicts his earlier statement that Galilei invariants are meaningful.

Hilbert went even further by demanding the covariance of all statements, basic and derived, theoretical and experimental, under the most general coordinate transformations, i.e., those occurring in general relativity. In fact he stated that "a proposition that is not invariant under every arbitrary transformation of the coordinate system must be regarded as *physically meaningless*" (Hilbert, 1924, p. 274). Equivalently: "a proposition is invariant and has therefore a physical meaning if it holds for every arbitrary coordinate system" (Hilbert, *op. cit.*, p. 278). As with the cases of Galilei and Lorentz invariance discussed a while ago, what is actually involved is a criterion of universal (or frame-independent, hence observer-free) *truth* not one of factual *meaning.*

Hilbert, and Weyl before him, may have been using the term 'meaning' in a colloquial sense for, taken literally, their sentences concerning objective meaning are meaningless. In any case they should not be construed as formal definitions of "meaning" as covariance. (See however Suppes, 1967.) If they were so construed then almost every physical statement would have to be ruled out as meaningless. Only the *basic* physical law statements, such as the variational principles and their immediate consequences, can be required to be covariant with respect to certain transformation groups: their solutions *must* be frame dependent (relative) if they are ever to be subjected to experimental tests, for test devices are in the habit of being attached to some frame or other and of yielding results that are seldom the same for every other frame (Bunge, 1961a, 1967b).

Examples like the following have been offered in support of the claim that covariance (of some kind) is required for objective meaningfulness. ⌜The car is at rest⌝ is not an invariant (or frame independent) statement, for the same car is really moving relative to almost every reference frame other than our planet. That statement is an instance of an incomplete or ill formed statement. The complete, well formed statement is ⌜The car is at rest in (or relative to) the ground⌝ (or the ferry or whatever may be the case). The latter statement is unobjectionable: it is meaningful and perhaps even true. But it is not generally covariant: it holds (is true) relative to one frame only. To disqualify ⌜The car is at rest⌝ we do not need a requirement as severe as general or even Lorentz or even Galilei covariance. Our criterion of meaningfulness in Sec. 2.3 will suffice, for the predicate "is at rest" does not belong to any body of contemporary factual knowledge: the concept of velocity, in particular zero velocity, involves the one of reference frame. Indeed, the velocity function is a certain function on the set of ordered triples: physical system p – reference frame f – velocity unit u. Hence while ⌜$V(p, f, u)=0$⌝ is well formed and meaningful, ⌜$V(p)=0$⌝ is not. This simple solution to the problem squares with actual scientific practice. And it is far more economical than replacing ordinary logic by some system of three valued logic (true, false, and meaningless) in order to accommodate freaks like ⌜$V(p)=0$⌝ or ⌜The mass of c equals 5⌝, as has actually been suggested (Suppes, 1959, 1965, 1967). As emphasized in Sec. 2.3, scientific constructs are not to be judged, let alone tampered with, in isolation from the theories they belong to, for the simple reason that only such theories exhibit their form and content. In particular scientific theory, not philosophy, is competent to determine whether or not (*a*) a given formula in scientific discourse is well formed and meaningful and (*b*) a given magnitude (or physical quantity) is frame independent or unit free or both (Bunge, 1971a).

In conclusion, invariance and covariance have nothing to do with meaningfulness, nor even with objectivity. Most of the known, meaningful, and even partially true formulas of physics are not generally covariant. Covariance, then, is not necessary for meaningfulness. Nor is it sufficient, as it is possible to propose any number of formulas that are covariant with respect to certain transformations but are devoid of factual meaning. Consequently covariance does not define meaningfulness. On the other hand covariance is necessary (though insufficient) for frame independent (hence observer free) truth, a condition only a few basic law statements satisfy. (See Bunge (1959b) and (1967b).)

6. CONCLUDING REMARKS

Our view on meaning and significance combines sense with reference. Thus a term such as 'electron' signifies both "the lightest material unit of electric charge" and any concrete instance of this predicate. It therefore consecrates and systematizes the ambiguity of the word 'meaning' in ordinary languages – without however conflating the two meaning components.

In our semantics there is no separate theory of meaning. But for the meaning calculus sketched in Sec. 2.1, our theses on meaning are mostly definitions and they make sense only against the background of the theories of sense and reference expounded in Part 1. Once the latter are assumed our definitions of meaning, meaning change, significance, synonymy, and their kin, look natural and even trivial. In point of fact our definition of meaning as sense together with reference cannot compete in boldness with any of its rivals: that meaning is verifiability, or truth condition, or use, or understanding, or information, or general covariance, or what not. We are proposing the cowardly view that meaning is just meaning.

So far we have been concerned with that which makes a construct what it is – namely its sense and its reference. Change either and a different construct will emerge. Not so truth and extension: these bear on fully meaningful constructs of certain kinds and, if factual, they are determined *ab extrinseco* rather than by analysis. In particular, whereas every factual statement is born with a sense and a reference, it is assigned no truth value until after having been tested for truth (Sec. 5.1). Moreover every truth value assignment is corrigible in this case: the same factual statement may now be assigned a truth value, now another. This is why we could not tackle the problem of truth before. But now the moment of truth has come.

CHAPTER 8

TRUTH

Logicians and mathematicians may be envied (or perhaps pitied) for needing a single concept of truth, namely that of formal truth. And they must be commended for having subjected it to a rigorous theory, namely Tarski's, now incorporated into model theory. Moreover, since in this theory the concept of (formal) truth is derivative (definable in terms of satisfaction in a model), formal scientists need not regard it as basic. Besides, this concept of truth raises no questions of confrontation with any factual facts: in logic and mathematics checking and proving are purely conceptual operations.

Factual scientists and the semanticists of factual science do not have it so easy. They have to contend with a radically different concept of truth, as suggested by the typical statement ⌜Theory T is an approximately true representation of factual domain F (or a good approximation in view of the evidence concerning F)⌝. What is at stake here is the concept of *partial truth of fact*, which – in contrast to that of complete formal truth – does not occur in the semantics of formal science. In matters of fact there is not only factual reference but also, and consequently, adequacy or inadequacy with respect to fact. Moreover such adequacy or inadequacy is rarely complete: unlike logical truth and logical falsity, factual truth and factual falsity are not polar opposites but contraries since, though incompatible, they are not exhaustive. This is sometimes not realized because science employs ordinary logic, hence seems to espouse a two valued theory of truth. But logic is the theory of deduction not the theory of truth. It is perfectly possible to process logically a set of statements that are neither fully true nor completely false: this is what science does. Hence we need a theory of partial truth of fact consistent with ordinary logic. In the present chapter we shall explore a chromatic (many valued) theory of truth matching with black-and-white logic.

1. KINDS OF TRUTH

1.1. *Truth Bearers*

Granting that "truth" makes sense at least in some contexts, the very first questions we must try to answer are: (i) what kind of objects are the bearers of truth and (ii) what kind of object is truth itself. There is of course a number of views on this matter. All of them agree that sentences are involved in these questions and that sentences are physical objects (strings of sounds or written marks) belonging to some language or other. But the various views assign sentences different roles in relation to truth. Table 8.1 exhibits the main traits of the principal views. (We have not included the "no truth" view of truth, set forth by Ramsey (1931), for being so glaringly at odds with the practice of testing for truth.)

The first view, naive idealism, is entrenched in the Western modes of speech and thinking. It seems right in holding truth to be a property of propositions (as distinct from judgments and sentences), wrong in regarding propositions as eternal objects which man can only "discover". But at least this phantasy of the autonomously existing propositions is sometimes heuristically fertile. Thus a mathematician may claim that a theorem that has not yet been formulated, let alone proved, is true – and he may set out to "discovering it". The obvious criticism is that nonexisting objects have no properties. Just as my unborn great-grandson is neither here nor there, so unthought formulas are in no theory (except potentially) and a fortiori they are neither true nor false. Moreover, existing statements that have not yet been proved are conjectures that should be assigned no truth values except for the sake of argument. Likewise a factual hypothesis cannot be said to have been true (or false) from all eternity, even before it was formulated. We can assign it a truth value only after having subjected it to tests – and even so the assignment may be provisional. But the proposition must have been stated to begin with.

In conclusion we reject naive idealism. We retain however the idea that constructs, e.g. propositions, are distinct from both thoughts and sentences, and that they can be true or false. But instead of assuming that there are propositions in themselves we assume that, to exist, propositions must be thought or uttered (or written) by some rational being in some language or other. And rather than claiming that truth and

TABLE 8.1

Main views on the nature of truth and its bearers

View	Sentences	Propositions	Truth
Naive idealism, e.g., Platonic "realism".	Express some propositions (the known ones).	Timeless objects existing by themselves	Intrinsic property of propositions. These are T or F whether we know it or not.
Neoidealism, e.g. Frege and the early Russell.	Express all propositions.	The meanings of sentences.	The object denoted by a true sentence.
Naive materialism, e.g. nominalism.	Physical objects with a factual reference and a truth value.	There are none.	Property of sentences in relation to their denotata.
Conceptualist materialism.	Express all propositions. Devoid of semantic properties except vicariously, i.e., via propositions.	Indispensable fictions – equivalence classes of certain thoughts.	Assignable to some propositions on the strength of others. Factual truth: relative, partial and ephemerous.

falsity are inborn, we assume that they can be attributed (in some degree) – as well as withdrawn. But we shall not elaborate on these psychological presuppositions: we shall take them for granted.

The second view, neoidealism, shares with Platonism the belief that there are constructs in themselves. But, unlike Platonism, it is muddled and even inconsistent. First truth and falsity are predicated of sentences (or of propositions, according to the way the ambiguous German *Satz* is translated). Next they are made into separate Platonic entities, *das Wahre* and *das Falsche.* Worse, truth value is equated with the referent, or nominatum, or designatum (*Bedeutung*) of the sentence (Frege, 1892). Thus the *Bedeutung* of ⌜Chimpanzees are clever⌝ will be said to be the same as that of $\ulcorner 2^2 = 4 \urcorner$, namely The Truth. This is not only a confusion but it renders the concept of reference unnecessary, as it makes all true sentences have the same "referent" (*Bedeutung*) regardless of their genuine referents. Frege could obviously afford to make a couple of mistakes. (Another mistake of his was to oppose the type of axiomatics initiated by Peano and Hilbert as well as the correlative notions of definition by postulates and conditional definition.) Unfortunately Frege's muddle has been taken seriously by some of the best logicians and philosophers of our time, such as Carnap (1947), Church (1951), Kemeny (1956) and Ajdukiewicz (1967b). (On the other hand Russell overcame this confusion.) We shall do well to keep clear from this muddle while retaining instead the distinction (made by Bolzano, Frege, Russell and Carnap) between sign and construct as well as Frege's notion of sense (*Sinn*) – which, unfortunately, he did not elaborate on. (Recall Ch. 4.)

The third view, naive (or vulgar) materialism, is best represented by Buridan, Hobbes, Hilbert, Tarski, and Quine. Like naive idealism it is simple, clear, consistent – and mistaken. A sentence, *qua* string of sounds or ink marks, is an object for physical science (e.g., acoustics or chemistry). A physical object becomes a linguistic object the minute it is regarded as a medium for expressing something. As the astronauts have discovered, in an uninhabited world there are no more sentences than there are propositions: at most there may be inscriptions – the bones of dead sentences. Further, sentences can be well formed or ill formed. If the former then they may but need not express any propositions. Thus 'Aleph zero took a bath' is a sentence that stands for no proposition: this is why it is devoid of significance. (If sentences were assigned truth values then, to make room for nonsignificant or nonsensical sentences, we should adopt some system of three valued logic – which, to say the least, would be unpractical.) Finally a metaphysical point: to attribute to strings of marks or sounds any semantic properties, such as meaning and truth, is to indulge in hylemorphism *à la* Plato or Husserl. Following Leibniz (1703) we assume then that propositions, not sentences, are the direct bearers of truth: sentences can only be vicariously true.

We are thus left with the fourth view, conceptualist materialism – a sort of materialist version of conventionalism. It can be summarized as follows. Propositions are a kind of construct and as such they are useful fictions: we do not claim that they exist by themselves but only that it is often convenient (for example in mathematics but not in metaphysics) to feign or pretend that they do. We do not assert that the Pythagorean theorem exists anywhere except in the world of phantasy called 'mathematics', a world that will go down with the last mathematician. And we do not pretend that unknown propositions exist but find it advantageous to proceed *as if* all the logical consequences of a proposition, both the few known ones and the infinitely many unknown ones, did exist in some conceptual context. (After all this is what the most fervent nominalist does when he equates an axiomatizable theory with the set of consequences of the axioms of the theory.) In this way we keep the heuristic and metatheoretical advantages of Platonism, which allows us to handle infinite sets of statements, only a tiny fraction of which will ever be formulated, even less justified. But we avoid the weird metaphysical hypothesis that every possible proposition actually exists in a ghostly Realm of Ideas.

Our view can be clarified by contrasting it to the prevailing doctrine of truth, namely Platonism. To a Platonist such as Frege every proposition exists from all eternity and has a truth value even if we do not know which one. This is one reason for defining a proposition as anything that is either true or false. In our semantics propositions are not so defined: they are characterized instead together with predicates, namely thus: a predicate is a function from objects of some kind to propositions. (Cf. Ch. 1, Sec. 1.3.) Once we have formed a proposition we may *find out* its meaning and *assign* it a truth value. Meanings are discovered by investigating both the reference and the context (e.g. theory) in which the propositions dwell. In any case propositions are born with a fixed meaning. There is no such thing as a meaningless proposition (in contrast to a sentence devoid of significance). True, it often takes a lot of work to discover the full meaning of a proposition: it requires displaying the full context and the logical relations in it.

On the other hand propositions are not born with a truth value: this one is assigned, if at all, once the proposition has been formulated. There are propositions that have not yet been assigned a truth value, e.g. because we have been unable to prove (or disprove) them or confirm (or infirm) them empirically. Surely, if p happens to be a theorem entailed by premises which have already been evaluated, then we *discover* the truth value of p by proving p. But not all propositions are provable. Axioms are not, nor are data. And if a proposition is not provable then there is no truth value to be discovered. If p is not a theorem then either we assign p a truth value (on some ground or just for the sake of argument) or we don't. If we do, our assignment may be criticized and exchanged for another: we shall say that the truth value of p in the light of evidence e (empirical or theoretical) is v, only to point out that a different body e' of evidence might suggest a different truth value v'. And if we refrain from assigning p a truth value, e.g. for want of means or of interest, then p remains a proposition though one without a truth value. Which ruins the Platonic conception of propositions.

1.2. *Truth Values: Acquired*

Being constructs, propositions are *designated* (or expressed) by linguistic objects of a certain category, namely sentences. And, being constructs, propositions *have* a sense and a reference – the latter either actual or hypothetical. In addition to having a sense and a reference, some (actually most) propositions *can be assigned* a truth value. And a few such potentially true or false propositions are *actually assigned* a definite truth value – a value that need not be unalterable. To put it negatively: Not every proposition can be assigned a truth value, nor is every truth value assignment final.

To be more explicit, the following kinds of proposition are assigned no truth value at least within certain contexts:

(i) the propositions that are not decidable in (with the sole resources of) a given theory have no truth value in this theory;

(ii) the propositions containing empty descriptions, such as "the perfect man", are neither true nor false – unless they assert or deny the existence of their imaginary referents;

(iii) the statements that are formulated, but are neither postulated nor proved nor confirmed nor made plausible, have no definite truth values. For example, ⌜$p \vee \neg p$⌝ in intuitionistic logic, and ⌜There are black holes⌝ in current astrophysics and cosmology.

Since not every statement can be assigned a truth value we shall construe the truth value assignment (function) $\mathscr{V}$ as a *partial* function, i.e. as a function on a proper subset S_D of the set S of all statements. (I.e., we adopt the truth value gap view.) We shall characterize $\mathscr{V}$ in Sec. 3. But before doing so we must discriminate among the various kinds of truth, for different kinds call for different value assignment procedures.

Figure 8.1 summarizes our views so far.

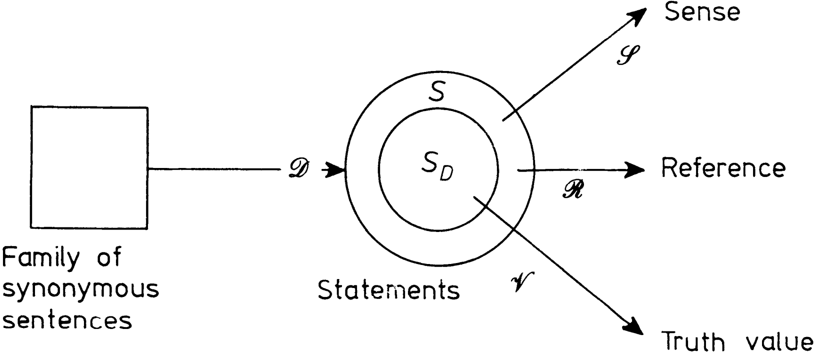


Fig. 8.1. Every set of synonymous sentences designates a statement in S. Every statement has a sense and a reference. Some statements are assigned a truth value.

1.3. *Quadruple Truth*

Consider the statements in Table 8.2, every one of which may be regarded as being true in its own context.

It would be nice if all these statements were true in the same manner, i.e., if a single concept of truth were applicable to them all. If this were the case then either the coherence or the correspondence doctrine might be able to span the whole set of statements. And if the former were the case, i.e., if every true statement were a *vérité de raison*, then model theory might suffice. That is, the semantics of mathematics would account not only for statements such as ⌜Formula ϕ is true in model M⌝ but also for propositions like ⌜The cost of living is rising steadily⌝. Unfortunately model theory does not help us with the second one nor with any statement in factual science, for the latter contains only fully interpreted formulas whereas model theory applies only if alternative interpretations (in mathematical structures – not with reference to the world) are possible. (Cf. Ch. 6, Sec. 2.) In particular, model theory is competent to handle formulas that hold under every interpretation of the variables involved – i.e. in all models. (These are the tautologies.) In sum model theory, which formalizes and systematizes the coherence view of truth, is not universal: it does not even apply to the whole of mathematics.

TABLE 8.2

Four kinds of truth: logical, mathematical, factual, and philosophical

Kind	Example
1 *Logical*	For any statement p, $\neg(p \mathbin{\&} \neg p)$
2a Abstract	In a Boolean algebra, for any element x, $x \wedge \bar{x} = 0$
Mathematical	
2b "Concrete"	In an algebra of sets, for any set S, $S \cap \bar{S} = \emptyset$
3a Theoretical	Let p_0 be the probability of allele A in the first generation and μ the mutation rate from allele A to allele a. Then the probability of the old allele A in the nth generation is $p_n = p_0(1-\mu)^n$
Factual	
3b Empirical	Almost all the individuals in the 100th generation of the bacillus X were observed to be mutants in the trait Y that had been observed in the fraction p_0 of the original population
4a Semantical	Meaning precedes truth
Philosophical	
4b Metaphysical	Every thing changes

On the other hand the correspondence theory of truth leaves us in the lurch with regard to logic and mathematics, which need not fit any facts in order to be valid. And it is unlikely that a third view could apply to all kinds of truth; at any rate no such all-encompassing theory seems to have been proposed. We had better recognize that the word 'truth' designates at least four different concepts, every one of which must be characterized in its own way: logical truth, mathematical truth, factual truth, and philosophical truth. Table 8.3 exhibits some of the salient peculiarities of mathematical truth and of factual truth. We shall take a

TABLE 8.3

Characteristics of three kinds of truth

Mathematical		Factual (e.g., biological)
Abstract (e.g., lattice theoretical)	"Concrete" (e.g., number theoretical)	
No formula has a truth value by itself, independent of an interpretation.	Most formulas have a truth value. Only the undecidable propositions lack a truth value (in the system in question).	Many formulas are assigned a truth value.
A relation between uninterpreted formulas and definite conceptual structures (models)	A relation between assumption and consequence.	A relation between statements and their referents.
Definable in terms of satisfaction (e.g., x satisfies ϕ in M) or of proof. Hence dispensable.	Definable in terms of provability in the case of syntactically complete theories. Hence dispensable in this case.	Not definable but characterizable. Indispensable.
Holds in a whole category of structures, e.g., the category of groups.	Holds only in a limited context, i.e., the theory of a model.	Holds only for a limited domain.
Can be total.	Can be total. Exception: approximation theory.	Can be total only in simple cases. Most theoretical factual statements are only partially true.
No decision procedure in most cases.	Usually decidable by proof or by counterexample.	No decision procedure. Only specific criteria for estimating truth values.
Truth valuation a purely rational procedure: experimenting infrequent and only heuristically valuable. Hence truth value independent of the world.		Truth valuation based partly on observation. Hence truth value contingent upon the world.

closer look at the former in Sec. 2.1, whereas the *vérités de fait* will be examined in Sec. 2.2.

2. TRUTH OF REASON AND TRUTH OF FACT

2.1. *Truth of Reason*

A truth of reason is, of course, one that can be established by reason alone. A number of kinds of truth of reason have been distinguished, among them the following.

(i) *Dictionary truth* or *veritas ex vi terminorum* – e.g., a nominal definition.

(ii) *Truth by request* or postulation – e.g., a postulate in a mathematical theory.

(iii) *Truth by proof or deduction* – e.g., a theorem in a mathematical theory.

(iv) *Logical truth* or *veritas ex vi formarum* or tautology – e.g., any valid formula in a given system of logic.

(v) *Truth by exemplification* or satisfaction in a model.

The first two hardly deserve to be classed among truths: dictionary "truths" are mere conventions; and mathematical postulates are proposed because they summarize theories not because they are assumed to be true in themselves. (If abstract they may be true in or relative to some model, if "concrete" they generate truths by deduction.) The third, truth by deduction, is another case of abuse of the word 'truth': if its peculiarity is that it is deducible from a set of assumptions then the concept of truth is redundant here. Only the last two concepts of truth are legitimate, namely those of true under all interpretations (logical truth) and true under some interpretations (mathematical truth). They are the object of model theory (Tarski, 1954–55; Robinson, 1963; Bell and Slomson, 1969; Chang and Keisler, 1973). We shall presently take a look at the model theoretic concept of truth – but only to better clarify, by way of contrast, the notion of factual truth.

Consider the formula $\phi = \ulcorner x^2 + x = 0 \urcorner$. This semiabstract formula is *satisfied* by the number -1, i.e. it becomes a *true* formula when x is interpreted as -1. Equivalently: the valuation $x := -1$ satisfies ϕ or produces a true statement. Also: ϕ is *satisfiable* in the structure $\mathscr{Z} = \langle Z, +, -, 0\rangle$, where Z is the set of integers. But the very same formula is also satisfied by the negative of any unit matrix I. Equivalently: ϕ is *satisfiable* in the ring $\mathscr{M}$ of square matrices. And so on and so forth. In symbols:

$$\models_{\mathscr{Z}} \phi(-1), \qquad \models_{\mathscr{M}} \phi(-I), \quad \text{etc.}$$

The notion of satisfaction can be extended to a set of formulas. For example, all of the axioms of the abstract theory of groups are satisfied by, or hold for, the integers. A proof that the structure $\mathscr{Z} = \langle Z, +, -, 0\rangle$ is or exemplifies a group $\langle S, \circ, ', e\rangle$ consists in showing that the valuation (interpretation)

$$S := Z, \qquad \circ := +, \qquad ' := -, \qquad e := 0$$

satisfies the axioms of group theory taken jointly. In other words, these abstract formulas (open formulas), when assigned the above interpretation, become formulas recognized as true in a specific (often familiar) field – in this case elementary number theory. The validity of the latter goes unquestioned in the proof. What one shows is that one conceptual structure (the abstract one) fits another conceptual structure (a specific or "concrete" one) – a clear case of *truth as coherence.*

This model theoretic concept of truth as satisfaction in a model serves to define the concept $\vdash$ of deducibility, namely thus. The formula ⌜If x and y are in L, then $x \wedge y \preccurlyeq x \vee y$⌝ can be said to *follow* from the axioms of lattice theory because, in *every example* of a lattice, meets precede joins. On the other hand a special theorem for chains, such as ⌜For any x and y in L, $x \preccurlyeq y$ or $y \preccurlyeq x$⌝, does not follow from the axioms for L. In general: let A be an abstract axiom system and let ϕ be a formula containing only concepts occurring in A. Then $A \vdash \phi$ iff every model of A is a model of ϕ, i.e. $A \models \phi$. This relation between $\vdash$ and $\models$ is easily generalized to a whole set S of abstract formulas. Call $\mathscr{M}(S)$ the whole lot of structures in which the formulas in the set S are satisfiable. Then the theory of $\mathscr{M}(S)$ is the set $\mathscr{C}n(S)$ of logical consequences of S. The notion of truth by proof is thus reduced, in a sense, to that of truth by exemplification or truth in a model. That this does not eliminate the methodological differences in proof types and procedures is beside the point: semantics is not concerned with methodological problems.

We have recalled succinctly the model theoretic concept of truth only to bring out its differences with the concept of factual truth employed in factual science. The most outstanding differences are these:

(i) The objects said to be true in some (extensional) model, or under some interpretation, are *abstract formulas.* What satisfies or fails to satisfy an abstract formula is always some construct or other. Because snow is not a mathematical object, it makes no sense in mathematics to state that snow satisfies the semiabstract formula $\ulcorner x$ is white$\urcorner$. Nor does it make sense in the context of factual (ordinary or scientific) knowledge, unless we are willing to assign a semantic property to a material thing.

(ii) An abstract formula, or propositional function, if satisfiable, and a fortiori if valid, is so *in one or more models.* By contrast, a factual statement need have no mathematical status at all: think of $\ulcorner$There are many beautiful children$\urcorner$. Moreover a factual statement, if true (to some extent), is true *of* the world not *in* a model. It would be ludicrous to write, say,

$\models_W$ Maxwell's electromagnetic equations, where W is the world.

if only because the real world is not a mathematically defined structure (an extensional model *à la* Tarski). The fashionable interpretation of models (or else of model sets in the sense of Hintikka (1969)) as possible worlds serves just one useful purpose – that of dressing Leibniz's notion of logical truth in modern garb. (To wit: p is logically true iff p holds under every interpretation of the nonlogical constants occurring in p, i.e. if p is satisfiable in every model – intuitively, if it holds in all conceivable, or conceptually possible, worlds.) This pseudo-ontological interpretation of model theory elucidates neither the concept of truth of fact nor that of real possibility. The former if only because truths of fact are more often than not partial, whereas a model set is constituted by formulas that are totally true under some interpretation. As to possibility, not any old model set describes a possible state of affairs or a possible course of events ("world"); only a set of law statements can do that. (The statements are at best approximately true when the possibilities are actualized. As long as the events are merely possible the corresponding statements do not form a model set.) Moreover, the translation of 'model' into 'possible world' has misled some philosophers into believing that, since logical truths hold (are valid) in every "possible world" (model or alternatively model set), and since the actual world is possible, logical truths must hold of reality, so that logic constitutes the basic ontology (Scholz, 1941) rather than a metaphysically neutral discipline. (More in Bunge, 1974a.)

To conclude. The model theoretic concept of truth elucidates the previously imprecise *coherence* theory of truth – contrary to Tarski's intention that it should formalize the correspondence theory. It also elucidates the notion of truth by proof. Consequently that semantic concept is of importance to pure mathematics. But it is irrelevant to factual knowledge, where we have not only coherence (or the fitting of constructs to one another) but also external reference.

Tarski's semantic conception of truth, in its mature (model theoretic) shape, revolutionized metamathematics. Moreover it drew the attention of some philosophers to a long neglected (or rather repressed) yet central problem of philosophy. But at the same time it persuaded some of the very best among them (notably Carnap, Popper, and Quine) that there was nothing problematic left about truth of fact. Yet Tarski's theory did not even attempt to grapple with the latter – so much so that all the technical work in model theory, starting with Tarski's pioneer papers (1936, 1954, 1955), is exclusively concerned with theories in pure mathematics – moreover with "formalized languages", i.e. abstract theories. Only an idealist could take model theory to be applicable to factual knowledge as well – since for him the world is a realization of an abstract idea. Let us then part company with model theory and take a fresh look at factual truth. (We shall return to Tarski's theory of truth in Sec. 2.4.)

2.2. *Truth of Fact: The Synthetic View*

A factual statement is one involving at least one factual predicate (see Ch. 6, Sec. 4.1). And factual truth (or falsity) is predicable of a factual statement in relation to some domain of facts and irrespective of its mathematical status.

Since the coherence theory of truth is not concerned with factual reference, it would seem that we must turn to the correspondence view of truth. According to the latter a statement is true if it fits the facts. Unfortunately the nature of this fitting has never been clarified: most of the time it is left in the twilight of metaphor, occasionally it is explained as isomorphism. Let us brush metaphors aside: they constitute no theory. Nor does the isomorphism thesis constitute a theory. To begin with isomorphism can only be defined between mathematically well defined structures – and reality is no such structure. Secondly, even the intuitive notion of isomorphism fails us here as shown by the facts that (*a*) every scientific theory contains constructs with no real counterpart and serving at the most as computational devices, and (*b*) every piece of reality ends up by showing traits that had not been contemplated by any theory. If there is correspondence between theory and facts it must be global not point by point. But it is doubtful whether this global correspondence can suffice to characterize factual truth. Indeed, consider the following cases, every one of which constitutes a counterexample to the (unformulated) correspondence theory of truth:

(i) ⌜There are no ghosts⌝ is factually true precisely because there is no such thing as a ghost.

(ii) Continuum mechanics is nearly true of most bodies – which are known not to be continuous.

(iii) The predictions and retrodictions calculated on theories assumed to be true refer to possible facts.

The reason for the failure to take the correspondence doctrine of truth beyond the metaphor phase and to have it surmount the indicated difficulties is both simple and radical, namely this. Statements cannot be compared or confronted with facts. Statements can only be confronted with other statements, and facts with other facts only. The expression 'Confronting a proposition with a fact' should be taken as short for 'Confronting a judgment (=brain process) with another fact' or, simpler, 'Thinking of a fact'. What holds for confrontation holds for fitting. A statement does not fit facts the way clothes fit people: it can only "fit" another statement, or "agree" with it barring certain details. In any case semantics is not equipped to investigate the mental processes of confrontation and adaptation of ideas to facts: it can only handle the confrontation among statements.

The following kinds of *inter-statement confrontation* are of particular interest to factual science:

$t-t'$ theoretical statements *vs.* theoretical statements
$t-e$ theoretical statements *vs.* empirical statements
$e-e'$ empirical statements *vs.* empirical statements.

Examples.

$t-t'$ The probability of a radiative transition of an atom from one energy level to another, calculated in nonrelativistic quantum theory, *vs.* the same probability calculated in relativistic quantum theory.

$t-e$ The probability of a radiative transition of an atom from one energy level to another, calculated in some theory, *vs.* the intensity of the radiation of a collection of atoms of the same kind, measured with a given equipment and technique.

$e-e'$ The intensity of the total radiation of a collection of atoms measured with a certain equipment and technique, *vs.* the values of the same magnitude measured with a different equipment and technique.

Any two statements can be put side by side. But if the aim is to estimate truth values then only statements with a *common meaning* (not necessarily identical) should be paired. In other words, for two statements to compete they must share some of their sense and some of their referents. (See Figure 8.2.) Thus while it may be fruitful to compare the values of a reaction time obtained by different methods, it would be foolish to compare either with the international price of sugar.

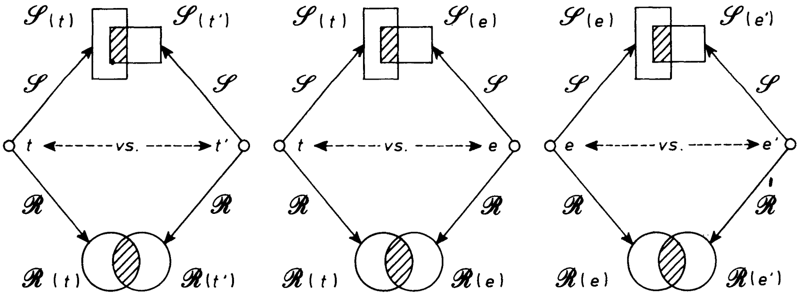

$\mathcal{S}(t)$
$\mathcal{S}(t')$
$\mathcal{S}$
$\mathcal{S}$
t ←-----vs.-----→ t'
$\mathcal{R}$
$\mathcal{R}$
$\mathcal{R}(t)$
$\mathcal{R}(t')$
$\mathcal{S}(t)$
$\mathcal{S}(e)$
$\mathcal{S}$
$\mathcal{S}$
t ←-----vs.-----→ e
$\mathcal{R}$
$\mathcal{R}$
$\mathcal{R}(t)$
$\mathcal{R}(e)$
$\mathcal{S}(e)$
$\mathcal{S}(e')$
$\mathcal{S}$
$\mathcal{S}$
e ←------vs.------→ e'
$\mathcal{R}$
$\mathcal{R}$
$\mathcal{R}(e)$
$\mathcal{R}(e')$

Fig. 8.2. Confronting scientific statements. Condition: both the senses and the reference classes must have nonempty intersections. Warning: $\mathcal{S}(e)$ and $\mathcal{S}(e')$ are rather vague.

The common meaning condition is most readily satisfied when the two statements are either theoretical or empirical. Difficulties arise when one of them is theoretical and the other empirical. And if the latter happens to be experiential (a sense datum), rather than observational or experimental, then there may be no point in confronting it with a physical object statement. For example, ⌜I am feeling hot⌝ differs both in sense and in referent from ⌜The present air temperature is 40 °C⌝. Hence neither can refute the other. In general, sense data are of little use in science precisely for this reason. Before a datum can become an item of evidence it must be depersonalized. But this is not enough: the datum must also be couched in the language of the theory it is supposed to be confronted with. For example, pointer readings must be interpreted as values of magnitudes handled by some theory. And before the latter can be confronted with any data it must be adjoined to special assumptions relevant to the situation that is being handled, as well as subjected to certain purely mathematical operations such as integration or substitution of reference frame. In short, the raw data are not comparable to the pure theory: the former must be lifted up to a theoretical level which is in turn lower than the axiom level. All this is a matter for the special sciences as well as for the general methodology of science (see e.g. Bunge, 1967a, Ch. 15, and 1973b, Ch. 10). It is also a matter for methodology to decide the conditions under which two comparable statements *agree* with each other, perhaps to within the preassigned experimental error. (See the next section.)

What is of overriding interest to semantics is (*a*) that the theory-fact confrontation boils down to confronting two sets of *statements*, and (*b*) that a theoretical statement is declared to be true if it "agrees" with some other *statements* (some of them empirical and others theoretical). It would seem that the coherence theory of truth does hold for factual science after all. It does but only in part: although a factual theory is true just in case it "agrees" with another set of propositions, all the parties involved – those in the dock as well as those in the jury box – happen to have a factual reference. We arrive thus at a sort of synthesis of the coherence and the correspondence views of factual truth – whence the name *synthetic* view. The coherence consists in inter-propositional agreement, the correspondence in factual reference. And the correspondence or degree of adequacy is tested by coherence: the latter supplies the criterion of truth not a definition of it. (Cf. Rescher, 1973.)

Another, no less important feature of factual truth of interest to semantics is this: it is *partial* more often than total – i.e. it comes in degrees. Figure 8.3 illustrates this contention (not contended by scientists though

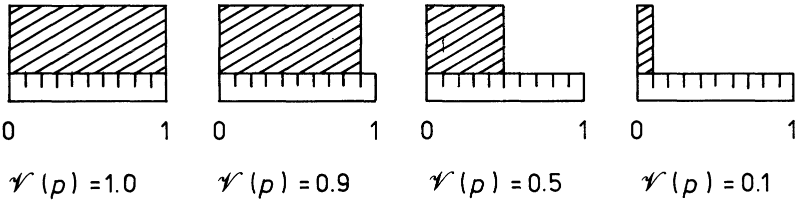


Fig. 8.3. Outcomes of a truth valuation of $p = \ulcorner$The box is 1 cm long$\urcorner$. (*a*) Complete truth; (*b*) approximate truth; (*c*) half truth; (*d*) near falsity.

brushed aside by logicians). The truth valuation function $\mathscr{V}$ occurring in the caption will be defined by Criterion 8.1 in Sec. 2.3. In the case of Figure 8.3 it is assumed that the ruler indicates its own actual length as well as that of the box. This hypothesis is the basis for assigning p a truth value.

2.3. *Truth Values: Conditional*

Although methodology is no substitute for semantics it may supply useful clues for investigating the semantics of science. In particular, if we wish to find out what factual truth is we shall find it useful to become acquainted with the way truth values are assigned in science. A very quick review will suffice for our purposes.

Whenever two statements, s and s', are confronted with a view to evaluating either of them, the following situations can arise:

(1) *Neither statement is taken for granted* (presupposed or assumed).

(*A*) *The two statements turn out to agree.* Unless there is a logical relation between them, both statements are assigned the same truth value: $\mathscr{V}(s) = \mathscr{V}(s')$. If s implies s' then $\mathscr{V}(s) \leqslant \mathscr{V}(s')$. In either case there is mutual confirmation but no independent assignment of truth values: for all we know the two statements could be equally false. (Moral: Confirmation, though necessary, is insufficient.)

(*B*) *The statements disagree.* They are assigned different truth values: $\mathscr{V}(s) \neq \mathscr{V}(s')$ – but we still do not know how much each is worth. (Moral: Refutation too is insufficient.)

Upshot: If neither of the two statements is assumed, at least provisionally, the other cannot be assigned a truth value either.

(2) *One of the statements, say s, is assumed* (unquestioned in the given investigation).

(A) *The two statements turn out to agree.* Both are assigned the same truth value, namely one or nearly so. Possible cases:
- (*a*) Both statements are theoretical. Theory confirms theory.
- (*b*) s is theoretical, s' experimental. Theory endorses experiment.
- (*c*) s is experimental, s' theoretical. Experiment supports theory.
- (*d*) Both statements are experimental. Experiment supports experiment.

(*B*) *The statements disagree.* The statement under examination is assigned a truth value below unity: $\mathscr{V}(s') < \mathscr{V}(s) = 1$. Possible cases:
- (*a*) Both statements are theoretical. Theory infirms theory.
- (*b*) s is theoretical, s' experimental. Theory indicts experiment.
- (*c*) s is experimental, s' theoretical. Experiment infirms theory.
- (*d*) Both statements are experimental. Experiment indicts experiment.

(*C*) *The statements agree in a certain region R but disagree in $\bar{R}$:* see Figure 8.4. The statement under examination is assigned a truth value dependent on the point x of the explored area: $\mathscr{V}(s', x) = 1 - \varepsilon(x)$, where $\varepsilon(x)$ is the discrepancy between s' and the base line s at the point x. Possible cases:
- (*a*) Both statements are theoretical. Theory confirms theory in region R, disconfirms it in region $\bar{R}$.
- (*b*) s is theoretical, s' experimental. Theory confirms experiment in region R, indicts it in $\bar{R}$.
- (*c*) s is experimental, s' theoretical. Experiment confirms theory in region R, infirms it in $\bar{R}$.

(*d*) Both statements are experimental. Experiment strengthens experiment in region R, weakens it in $\bar{R}$.

Upshot: If one of the two statements is assumed, if only provisionally, the truth value of the other can be estimated or at least bounded.

We conclude that factual truth values are *conditional* or *relative* not absolute. Hence strictly speaking we should always write '$\mathscr{V}(s' \mid s)$', read 'the truth value of s' given, or assuming, s', instead of '$\mathscr{V}(s')$'. And, since there are only relative factual truth values in science, there would be no point in trying to analyze $V(s' \mid s)$ into the absolute truth values $V(s')$ and $V(s)$ similarly to the way conditional probabilities are analyzable into absolute ones. Which suggests that truth values cannot be probabilities. (More on this in Sec. 5.1.)

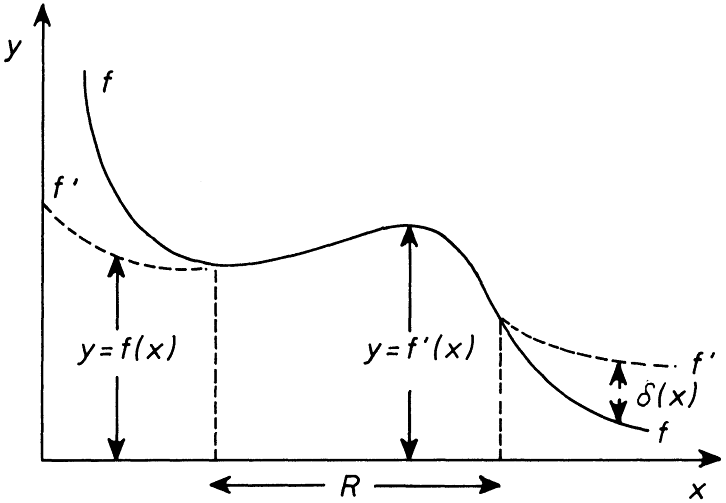


Fig. 8.4. A common situation in science: two statements, $s=(y=f(x))$ and $s'=(y=f'(x))$, agree in region R but disagree elsewhere. The relative discrepancy or error $\varepsilon(x)$ depends on the point x of the explored area and is proportional to the absolute discrepancy δ (x).

In general we shall not encumber our formulas with a symbol for the statement serving as a yardstick for the estimation of the truth value of another statement. But it will be convenient, nay indispensable, to indicate the baseline, in some way or other, whenever actual truth value assignments are made. A typical case in science is the evaluation, by alternative methods, of some magnitude M concerning an object b – for example the relative dilation of a metal bar. Suppose we have the following confrontation:

$$s = \ulcorner M(b) = m \urcorner \quad vs. \quad s' = \ulcorner M(b) = m' \urcorner,$$

where m and m' are rival numerical values obtained for the M-ness of b. (We are brushing aside, for being irrelevant to our purpose, the M-units as well as the experimental error.) The absolute value of the numerical discrepancy between the two statements is $|m' - m|$: this is the *error* made in accepting s' instead of s. If the error is small, then the truth value of s' will be close to that of s, i.e. near unity. But as the error increases the truth value of s' approaches zero. In general this error will depend on the referent b and its condition; hence the relative truth values themselves will exhibit this dependence.

The preceding considerations suggest the following

CRITERION 8.1 Let M and M' be two comparable functional representations of a given property of an object b, and let $s = \ulcorner M(b) = m \urcorner$ and $s' = \ulcorner M'(b) = m' \urcorner$ be estimates of M and M' respectively at b. Then the *relative truth value* of s' given (assuming) s equals

$$\mathscr{V}(s' \mid s) = 1 - \left| \frac{m - m'}{\mathrm{Max}\{m, m'\}} \right|$$

If the error is small, $\mathscr{V}(s' \mid s)$ comes close to 1; if the error is large, the truth value is near 0. *Example:* Let us compare the truth value of Boyle's law (B) relative to the law of Boyle and Mariotte (M), i.e., making the pretense that the latter is true. Since the statements are

$$M = \ulcorner p = aT/v \urcorner, \qquad B = \ulcorner p = b/v \urcorner,$$

we have

$$\mathscr{V}(B \mid M) = 1 - \frac{aT/v - b/v}{\mathrm{Max}\left\{\dfrac{aT}{v}, \dfrac{b}{v}\right\}} = 1 - \left| \frac{aT - b}{\mathrm{Max}\{aT, b\}} \right|.$$

Both for very low and very high absolute temperatures $\mathscr{V}(B \mid M)$ is near zero; it comes near unity for medium temperatures and certain gases. That truth values should depend on the kind of material and on the range of the physical variables involved may sound outrageous to both pure logicians and Platonists. But it is no news to a scientist, who is used to thinking of (factual) truth as holding to some extent *of*, or *with regard to*, definite external referents in definite conditions. This is precisely the peculiarity of factual truth: that it concerns facts. For this reason, as will be seen in the next subsection, the semanticist has no say in setting up truth conditions for factual statements.

Criterion 1 applies, in particular, to the comparisons of theoretical or calculated values with experimental values. In this case we have pairs of propositions such as these:

Theory $M(b)=m_t$, with m_t a knowable number

Experiment $M'(b)=m_e \pm \varepsilon_e$, with m_e and ε_e knowable numbers (m_e is the measured value and ε_e the random error characteristic of the particular run of measurements leading to m_e).

Unless the possible values of M are well spaced, the chances are that the theoretical value m_t and the central experimental value m_e differ from one another – and both from the real value. We need then definite criteria allowing us to make definite choices. One such criterion, actually used if not formulated explicitly in the sciences, is

CRITERION 8.2 Let m_t and m_e be a theoretical and an experimental estimate, respectively, of a magnitude M representing a property of a thing b. Then

(i) $\ulcorner M(b)=m_t \urcorner$ and $\ulcorner M(b)=m_e \urcorner$ are equivalent, to within the error ε_e, iff $|m_t - m_e| < \varepsilon_e$;

(ii) the "true" (or "real") value of M for b is close to m_t iff (*a*) m_t coincides with m_e to within ε_e (i.e. if (i) holds) and (*b*) $\varepsilon_e/m_e \ll 1$.

2.4. *Truth Conditions*

A truth condition (or criterion) for a set S of formulas is a metastatement stipulating the conditions under which the members of S are (fully) true. Example:

If x is a real number then $\ulcorner x^2 > x \urcorner$ is *true* iff $|x| > 1$ or, equivalently,

$\ulcorner x^2 > x \urcorner$ is *satisfied* by all real numbers x such that $|x| > 1$.

Tarski's theory of truth (Tarski, 1936), based on the concept of satisfaction, provides a general truth condition for all the *abstract* formulas of logic and mathematics and it has become part and parcel of model theory (see e.g. Hermes, 1963). Some philosophers believe that the same trick works for all statements, in particular for factual statements. (And it certainly should if, as Tarski and Quine hold, there is no radical difference between the factual and the formal nor, a fortiori, between the synthetic and the analytic.) Thus Carnap illustrates the theory by contriving a toy semantic system for an object language with 7 specific (extralogical) signs: the individual constants x_1, x_2 and x_3, the predicate letters P_1 and P_2, and the left and right parentheses (Carnap, 1942, pp. 22ff.). The atomic sentences are all of the form "$P_i(x_j)$" with $i = 1, 2$ and $j = 1, 2, 3$. There is one designation rule for every specific sign, in particular the following:

'x_1' designates Chicago.

'P_1' designates the property of being large.

The truth condition for the sentences of the microlanguage is

The sentence '$P_i(x_j)$' is true iff the designatum of x_j has the designatum of P_i.

An instance of this truth condition is, of course,

'$P_1(x_1)$' is true iff Chicago is large.

It would be nice if Tarski's theory of truth were a stretch theory fitting all kinds of statements, whether formal or factual. But it is not. First, in the body of factual knowledge there are no *uninterpreted* formulas, such as '$P_i(x_j)$', containing predicates with no fixed sense. Hence the possibility of having factual formulas satisfied in alternative models is not forthcoming – whence model theory is irrelevant to our concern.

Second, to say that Chicago *satisfies* the open formula $\ulcorner x$ is large$\urcorner$ involves assigning a nonphysical (namely a semantic) property to a physical entity – which is unacceptable to a non-Platonist. Things are the other way around: $\ulcorner x$ is large$\urcorner$ has the (semantic) property that, when the dummy x is replaced by a name for Chicago it becomes a true proposition.

Third, there is no *universal* truth condition for all of the (interpreted) formulas with a factual reference, from ⌜Chicago is large⌝ to the sophisticated equations of mathematical biophysics. There are at most regional truth conditions such as

> ⌜The city of x is large⌝ is *true* iff the population of x is over one million people.

In this simple case the truth condition proves to be a trivial convention of demography – one that is neither portable nor eternal nor stipulated by semantics. In other cases, *if* there are truth conditions they are more complex. And in any event they are laid down (and upset) by the competent discipline. It should be obvious why this is so: a factual truth condition depends on the specific meaning (sense and reference) of the corresponding statement as well as on the possible procedures for putting it to empirical tests. Consequently factual truth conditions cannot be cooked up by semantics.

Fourth, in factual science there are hardly any truth conditions having the neat form of biconditionals of the form: ⌜*A*⌝ *is true iff B*, a generalization of Tarski's principle: ⌜*A*⌝ *is true iff A*. The most we get in factual science are conditionals such as ⌜If the theory (or hypothesis) T is true then the effect e is observable⌝. (Or, equivalently, the corresponding counterfactual sentence 'If T were true then e would be observable.') But such conditionals work as clues for guessing at truth values not as criteria for assigning definite truth values. Indeed, the validation of the consequent e confirms the antecedent T without verifying it: in principle infinitely many alternative constructs T', T'',... might be substituted for T. It is only a whole lot of confirmations together with the compatibility of T with previously corroborated theories that allows one to (tentatively) assign (approximate) truth values to T. In sum, the normal situation in factual science is the lack of neat truth conditions and the presence of entire batteries of tests for (partial) truth (Bunge, 1967a, Vol. II, Ch. 15). And such assignments of degrees of truth are rarely final. (See Sec. 4.4.)

Fifth and lastly, since factual truth is rarely total, the truth conditions found in the usual treatments of mathematical logic are often inapplicable to it. A half truth such as ⌜Aristotle was a Chaldean philosopher⌝ fits neither of the standard truth conditions ⌜*A & B is true iff A is true and B is true*⌝, and ⌜*¬A is true iff A is not true*⌝. (Roughly, since one of the conjuncts is totally true and the other totally false, A & B is worth $\frac{1}{2}$ and so is its negate.) Hence logic, indispensable as it is for the control of inference, is utterly impotent to guide our assignment of factual truth values.

Instead of uniform and unchanging truth conditions, what we do meet in factual science are three kinds of conditions, all of them regional or subject matter dependent. We shall introduce them by way of example. Consider Galilei's law of falling bodies. There is, first, the *applicability condition* indicating the referents and the state they are in – for example a body immersed in a homogeneous gravitational field and in free fall in it. This condition occurs as the antecedent of the law statement: ⌜If a body falls freely in a constant gravitational field in vacuum, then G⌝, where $G = \ulcorner v(t) = gt + v_0 \urcorner$. The antecedent is not necessary for the truth of G but for its applicability or relevance: if the condition fails to hold the conditional is true but pointless. And the applicability criterion is *intra*theoretical: it owes nothing to test conditions. Rather on the contrary, testing presupposes that the applicability condition holds.

The second condition may be called an *ontological truth condition*, for it points to those referents that behave actually as indicated by the statement in question. In other words, the ontological truth condition for a statement is learned from experience and it exhibits the extension of the given formula. For example, the extension of G above is the collection of triples ⟨middle sized body, weak constant gravitational field, short time of fall⟩. This condition is *extra*theoretical and a posteriori. However, it is not theory free, for it sums up the outcome of tests conducted in the light of other (instrumental) theories and it evaluates the performance of the statement concerned in relation to (actual or possible) competitors.

A third and last metastatement of interest specifies the methodological conditions under which the given statement may be regarded as approximately true. It is, like the former, extratheoretical but, instead of specifying the kind of thing for which the statement holds (approximately), it concerns the particular empirical techniques employed in assaying it: this condition may therefore be called an *epistemic truth condition.* One such condition, in our example, would be this: ⌜G holds to within 1% for steel balls, in air, at sea level, and for distances of the order of 10 m, when tested with a sports chronometer and a commercial tape⌝.

All three conditions are special or subject matter dependent. Consequently they cannot be laid down by semantics. The semanticist may of course study them on condition that he leaves the methodological aspect to methodology and concentrates on the alethic aspect. But for this he needs a theory of degrees of factual truth, one elucidating the intuitive notion employed in factual science. We turn to one such theory.

3. DEGREES OF TRUTH

3.1. *The Problem and How to Fail to Solve It*

The notion of degree of truth, and the related notion of approximate truth, are employed everywhere in applied mathematics and in factual science. Examples: (*a*) most values of nonalgebraic functions, such as *log* and *sin*, are known only approximately; (*b*) all nontrivial measurement results are approximate; (*c*) all theoretical statements are, at best, good approximations – and we hope to be able to improve on them. Provided certain assumptions are made, i.e., provided some statements are taken to be wholly true, one can often estimate the goodness of the approximation, i.e., the deviation from the truth. In particular, one may compute (*a*) differences between values given by rival theories, (*b*) discrepancies between theoretical and empirical values, and (*c*) random errors of measurement. Mathematical statistics, by elucidating several concepts of error and computing their probabilities, helps us to estimate the deviation from the truth and, in this way, the degree of truth. On the other hand neither the theory of probability nor mathematical statistics have rules for assigning probabilities to either hypotheses or data: this assignment has always been a philosophers' game.

Since applied mathematics and factual science are shot through with the concept of approximate truth, it is a task for the semantics of science to elucidate this concept, i.e., to propose theories of degrees of truth matching scientific practice. There have been several attempts to do so. They fall into four main classes: (i) many valued logics (e.g. Moisil, 1972); (ii) the semantic interpretation of probability, i.e. the equating of probabilities and degrees of truth (e.g., Łukasiewicz, 1913; Reichenbach, 1949); (iii) Popper's theory of verisimilitude (Popper, 1963b), and (iv) the author's former theory of partial truth (Bunge, 1963a). None of these can be counted as successful, although the last may come close to providing a realistic explication of the concept of degree of truth, if only because it does not involve the mythical concept of probability of a statement.

Knowing why those attempts failed may help in avoiding similar mistakes. The reasons for those failures are, in a nutshell, the following. The systems of many valued logic may be mathematically interesting and they have exerted a liberating influence by showing that ordinary logic is neither logically necessary nor psychologically compelling. But none of them has attained the maturity needed to cope with actual deductive inference as practised in mathematics. None of them has ever been used in mathematics or in science, where ordinary logic is perfectly adequate. Moreover, none of them is likely to be used. First, because changing the logic in any one field of research would require changing it in all related fields: the revolution would have to spread all over mathematics and science in order to allow for contacts among theories. Second, it would be imprudent to relax the standards of criticism (Popper, 1970). Third, the main rationale for many valued logics is a mistake, namely the belief that logic should be a theory of truth. Logic must be regarded instead, in line with the Aristotelian tradition, as a theory of deduction not as a theory of truth. We should then be able to keep the calculi of ordinary logic even if we intend to adopt a many valued theory of truth. This policy will be implemented in the next subsection.

As to the probability theories of truth, they all rest on the mistaken assumption that there *are* ways of assigning probabilities to statements. In point of fact, whereas scientists often succeed in calculating and measuring probabilities of certain *facts*, e.g., events, no one has ever proposed a general procedure (other than arbitrary fiat or else betting) for assigning any probability values to *statements*. Mathematics would have nothing against it, for the set of statements qualifies as the support of a probability measure. But it so happens that there *are* no rules for assigning numerical values to such probabilities, as a consequence of which no one has ever succeeded in estimating the probability of any given factual statement. This difficulty alone disqualifies all the theories of truth degrees based on probability, whether they equate truth with probability or with some function of the latter. (More on this in Sec. 4.2.) On the other hand there are more or less definite rules for estimating relative degrees of truth, as we saw in Sec. 2.3.

Finally the author's previous theory of partial truth does not hinge on the concept of probability and it leaves ordinary logic intact. Besides, it incorporates the notion of discrepancy or error as it occurs in the theory of errors. However, it has some serious defects pointed out by the author himself (1963a) and by some readers. For one thing it overrates confirmation. For another its truth function is discontinuous. Thirdly, its multiplication theorem (or rather axiom) is so complicated that it is nearly impossible to compute by hand the truth value of a conjunction with a reasonable number of conjuncts. We shall presently expound an alternative theory of degrees of truth that shares the virtues but not the defects of the former theory.

The new theory to be presented is based explicitly on an *alethically neutral conception of logic* – one that can be attached to any number of alternative theories of truth. This construal of logic is nothing but an explicitation of Bolzano's remark that we should distinguish a proposition from the statement (actually a metastatement) that the proposition is true. Briefly, $p \neq \ulcorner \mathscr{V}(p) = 1 \urcorner$. One advantage of the neutral conception of (formal) logic is that it allows us to talk about partial truth and moreover to hitch any theory of partial truth to the calculi of ordinary logic. Scientists, though perhaps not formal logicians, will appreciate this advantage. A second advantage of this truth-free conception of logic is that it permits us to employ the dialectical method – in Parmenides' not in Hegel's sense. Indeed, the dialectical method, universally employed in mathematics and in science, consists roughly in exploring the consequences of an assumption *before* evaluating it and *in order to* evaluate it. To carry out this method we must assume that the propositions under examination obey the laws of logic whether or not they turn out to be true. The most we might need is the pretense that statements are true or false whether we know it or not. But we take this as a make-believe useful for heuristic purposes: we assume explicitly only that (some) statements *can be* assigned truth values, or rather truth degrees, not that they are born with an intrinsic and everlasting truth value.

Compare this formal or alethically neutral conception of logic with the alethic interpretations. Among these, the standard view is the model theoretic or referential one, which employs the notions of satisfaction and of (formal) truth, e.g., in setting up truth conditions such as truth tables. An alternative view that is receiving some attention these days is the so called substitution interpretation (Barcan Marcus, 1962). According to this view ‘$(\exists x)\, Px$’ should be interpreted as “Some substitution instance of Px is true” – and similarly for universal statements. Either view may be adequate for logic but, by the same token, it does not seem suitable for the applications of logic. For one thing they render the dialectical method inapplicable, insofar as they demand that every proposition be true or false from birth, whether we know it or not: i.e. it does not make room for tentative truth value assignments. For another, the alethic conceptions of logic employ a single concept of truth and moreover the one of total truth. If applied to factual science, where evidence is never complete and final, the alethic conceptions of logic may lead to the aberration of thinking that, precisely for those shortcomings, factual science need not abide by ordinary (classical) logic. (For a candid statement of this dangerous thesis see Birkhoff, 1961, Ch. XII.)

But enough of criticism: let us see how we can merge two valued logic with the idea that truth, unless it is formal, comes in any number of shades.

3.2. *Axioms*

Our theory of degrees of truth will treat complete truth and complete falsity as the two end points of a whole gamut of truth values, while retaining all the algebraic features of ordinary logic, for the latter is built into mathematics and science. In other words we shall (*a*) assume logic to be given in a purely syntactic fashion rather than with the help of (formal) truth values, and (*b*) join to it a valuation function $\mathscr{V}$ with values in a numerical interval, say the unit interval of the real line.

One way of implementing this programme is as follows. Consider the set S of all statements in a given field of inquiry, such as a scientific theory. Group all the statements in S that are logically equivalent to each other. That is, form the equivalence class $[s]$ of every statement s in S under the relation of logical equivalence: $[s]=\{s'\in S \mid \ulcorner s'\Leftrightarrow s\urcorner \text{ is a tautology}\}$. Call $[S]$ the set of all such equivalence classes, i.e. the quotient of S by the relation of tautological equivalence. It is well known that $[S]$ has the lattice structure. (This holds also for any extension of a given S but it need not hold for the union of arbitrary sets of statements, as they may be mutually incompatible. Thus the union of classical mechanics and quantum mechanics does not possess the lattice structure.) Moreover $[S]$ is a complete complemented and distributive lattice with null and unit elements – in short a Boolean algebra. The Boolean operations on the set $[S]$ of equivalence classes are defined in logical terms as follows: for any p, q, r in S,

$$\overline{[q]} = [p] \text{ iff } \ulcorner p \Leftrightarrow q \urcorner \text{ is a tautology}$$
$$[q] \cup [r] = [p] \text{ iff } \ulcorner p \Leftrightarrow q \vee r \urcorner \text{ is a tautology}$$
$$[q] \cap [r] = [p] \text{ iff } \ulcorner p \Leftrightarrow q \,\&\, r \urcorner \text{ is a tautology.}$$

Likewise the least element $\square$ and the last element ▯ of the Boolean algebra of the equivalence classes of statements are defined by

$$\square = \{p \in S \mid p \Leftrightarrow q \,\&\, \neg q \text{ and } q \in S\}$$
$$\text{▯} = \{p \in S \mid p \Leftrightarrow q \vee \neg q \text{ and } q \in S\}.$$

(It is true that the algebra of quantification is far more complicated than this. However we need not go into it if our purpose is restricted to computing the truth value of truth functional compounds in terms of the truth values of their components. Nor need we stop at the possible objection that the formulas of quantum mechanics fail to constitute a distributive lattice. This opinion is false: suffice it to recall that the quantum theories, just as any other scientific theories, involve only classical mathematics, which has classical logic in its bones. Cf. Bunge (1967b) and Fine (1968).)

We now assume that there is a real valued function $\mathscr{V}$, defined on a certain subset S_D of S, such that

$$\text{for every } p \text{ and } q \text{ in } S_D,\ \mathscr{V}(p \,\&\, q) + \mathscr{V}(p \vee q) = \mathscr{V}(p) + \mathscr{V}(q).$$

The statements in the complement $S - S_D$ have no truth value because none can be assigned to them. In this residual subset we find the statements in S that cannot be proved with the sole resources of S, as well as the statements containing empty descriptions, such as "The perfect man does not exist" and "The massless body is not affected by gravity". (See Ch. 9, Sec. 2.) In addition to the above conditions on the partial function $\mathscr{V}$ we shall postulate that $\mathscr{V}$ assigns contradictions the lowest truth value, namely 0, and tautologies the highest, namely 1. We shall also check whether our theory gives reasonable results in typical cases of scientific inference.

The foregoing is spelt out in an axiom disguised as

DEFINITION 8.1 The structure $\langle S, S_D, [S], \square, \text{▯}, \cup, \cap, {}^{-}, \mathscr{V}\rangle$, where S is a nonempty set, S_D a subset of S, $[S]$ the quotient of S by the relation $\Leftrightarrow$ of logical equivalence, $\square$ and ▯ distinguished elements of $[S]$, $\cup$ and $\cap$ binary operations on $[S]$, ${}^{-}$ a unary operation on $[S]$, and $\mathscr{V}$ a function on S_D, is called a *metric Boolean algebra of statements* iff

(i) the structure $\langle [S], \square, \boxed{\,}, \cup, \cap, {}^{-} \rangle$ is a Boolean algebra, i.e. a complemented and distributive lattice with null element $\square$ and universal element $\boxed{\,}$;

(ii) $\mathscr{V}$ is a real valued function on $S_D \subset S$ such that, for any elements p and q of S_D,

$$
\begin{aligned}
&(a)\ \mathscr{V}(p \,\&\, q) + \mathscr{V}(p \vee q) = \mathscr{V}(p) + \mathscr{V}(q);\\
&(b)\ \mathscr{V}(p) = 0 \quad \text{for all} \quad p \in \square;\\
&(c)\ \mathscr{V}(p) = 1 \quad \text{for all} \quad p \in \boxed{\,}.
\end{aligned}
$$

Condition (*a*) is common to all metric lattices or measure algebras. Conditions (*b*) and (*c*) determine the range of the values of $\mathscr{V}$. These three conditions are jointly necessary and sufficient to determine $\mathscr{V}$. However, they are not enough to compute the value of an arbitrary propositional compound from the values of its components. (In other words, condition (*a*) is not a full multiplication theorem.) However, we shall see in Sec. 3.4 that this partial indeterminacy is not a serious practical drawback.

Before going any further let us give two warnings. First, contrary to appearances, $\mathscr{V}$ is not a probability measure on S_D, if only because S_D is a set of individuals not a field of sets (a σ-algebra) – as it should if it were to qualify as a probability measure. More on this in Sec. 3.6 point ix. Second, we are dealing with closed bodies of knowledge rather than arbitrary sets of statements – let alone the totality of factual statements. The reason for the restriction is this. Every scientific theory, if consistent, is an ultrafilter (recall Ch. 5, Sec. 3.1). But not all scientific theories in use at a given time are mutually consistent. Hence their union does not constitute an ultrafilter. (Moreover: in general the union of two theories is not a theory.) In other words, Boolean algebra reigns within every theory but does not govern the totality of scientific propositions, not even in a given field of research.

3.3. *Topologies of S_D*

We shall presently show that S_D has two topologies of interest to semantics, given by so many metrics.

DEFINITION 8.2 The function $\delta_{-} : S_D \times S_D \to [0, 1]$ which assigns to each pair of propositions $p, q \in S_D$ a real number between 0 and 1, such that

$$\delta_{-}(p, q)=|\mathscr{V}(p)-\mathscr{V}(q)|,$$

is called the *horizontal distance*.

This name for δ_{-} is not metaphorical, for δ_{-} has in fact almost all the properties of a distance function, as shown by

THEOREM 8.1 The structure $\mathscr{M}_{-}=\langle S_D, \delta_{-}\rangle$ is a quasimetric space, i.e. the distance function δ_{-} satisfies the following axioms:

(i) $\delta_{-}(p, q)=\delta_{-}(q, p)$
(ii) $\delta_{-}(p, q)+\delta_{-}(q, r)\geqslant\delta_{-}(p, r)$
(iii) $\delta_{-}(p, q)=0$ iff $\mathscr{V}(p)=\mathscr{V}(q)$

for any p, q, and r in S_D.

The quasi metric δ_{-} defines a topology in the space S_D. An open ε-neighborhood of $p\in S_D$ is the set

$$U_\varepsilon(p)=\{q\in S_D \mid |\mathscr{V}(p)-\mathscr{V}(q)|<\varepsilon\} \quad \text{with} \quad 0\leqslant\varepsilon\leqslant 1.$$

This is the set of statements that are equivalent to the given statement to within the tolerance (error) ε. For example, the set of possible confirmers q of a hypothesis p is included in the ε-neighborhood of p. See Figure 8.5.

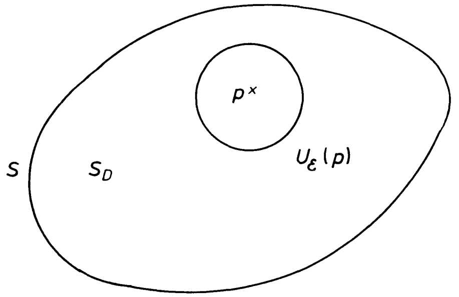


Fig. 8.5. An open neighborhood of $p\in S_D$. All the statements that agree with p to within ε are in $U_\varepsilon(p)$.

We can now formalize the notion of agreement between two statements, which we used in Sec. 2:

DEFINITION 8.3 For any given statement p in S_D and any preassigned real number $0\leqslant\varepsilon\leqslant 1$, the statement $q\in S_D$ *agrees with* p *to within* ε iff q is in the ε-neighborhood of p.

COROLLARY 8.2 Equivalent statements agree with one another.

Proof. By Theorem 2(iii), the distance between equivalent statements is nil.

Another equally natural topology is determined by a second distance function introduced by the following

DEFINITION 8.4 The function $\delta_{|}: S_D \times S_D \to [0, 1]$ which to each pair of propositions p, $q \in S_D$ assigns a real number lying between 0 and 1, such that

$$\delta_{|}(p, q) = |\mathscr{V}(p \vee q) - \mathscr{V}(p \,\&\, q)|,$$

is called the *vertical distance*. See Figure 8.6.

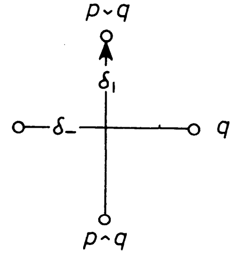


Fig. 8.6. Horizontal and vertical distances between propositions.

This other function deserves its name, as shown by

THEOREM 8.2 The structure $\mathscr{M} = \langle S_D, \delta_{|}\rangle$ is a quasimetric space.

This new metric defines a second topology in S_D. An open ε-neighborhood of $p \in S_D$ is now

$$U_\varepsilon(p) = \{q \in S_D \mid \delta_{|}(p, q) < \varepsilon\} \quad \text{with} \quad 0 \leqslant \varepsilon \leqslant 1.$$

The two truth spaces, $\langle S_D, \delta_{-}\rangle$ and $\langle S_D, \delta_{|}\rangle$, are separable (Hausdorff) because they are quasimetric. That is, for any two propositions p and q in S_D, there are open sets G and H in S_D such that p is in G and q is in H, and G and H are disjoint. Beyond their shared separability the two truth spaces are quite different, as shown by

THEOREM 8.3 The vertical distance between any two propositions is greater than or equal to their horizontal separation:

$$\text{If} \quad p, q \in S_D, \quad \text{then} \quad \delta_{|}(p, q) \geqslant \delta_{-}(p, q).$$

Consequently, any ε-neighborhood built with $\delta_{|}$ includes the corresponding set constructed with δ_{-}. Therefore the topology $T_{|}$ generated by $\delta_{|}$ is stronger than the topology T_{-} determined by δ_{-}. Because the vertical distance between a statement and its negate is maximal (i.e., $\delta_{|}(p, \neg p) = 1$), which is not the case with the horizontal distance, $T_{|}$ may be preferred to T_{-}. This is as far as we go in exploring the topologies for S_D determined by our truth valuation.

3.4. *Comparing Truth Values*

Let us now derive another few consequences from our assumptions. To this end we shall make free use of ordinary logic. And we shall bear in mind that stating a proposition is no indication of its truth value: the latter, if assigned at all, must be assigned by way of an extra metastatement such as, e.g., $\ulcorner \mathscr{V}(p) = \frac{1}{4}. \urcorner$

THEOREM 8.4 For any $p, q \in S_D$, $\mathscr{V}(p \vee q) \geqslant \mathscr{V}(p \,\&\, q)$.
Proof. By Definition 4 and Theorem 2.

THEOREM 8.5 For any $p \in S_D$, $\mathscr{V}(\neg p) = 1 - \mathscr{V}(p)$.
Proof. Set $q = \neg p$ in Definition 1.

THEOREM 8.6 The truth value of the antecedent of a wholly true conditional does not exceed the degree of truth of its consequent:
For any $p, q \in S_D$, if $\mathscr{V}(p \Rightarrow q) = 1$, then $\mathscr{V}(p) \leqslant V(q)$.
Proof. Set $\mathscr{V}(\neg p \vee q) = 1$ in Definition 1 and use Theorem 5 to obtain

$$\mathscr{V}(p) = \mathscr{V}(q) - \mathscr{V}(\neg p \,\&\, q) \leqslant \mathscr{V}(q).$$

COROLLARY 8.3 Equivalent statements in S_D have the same truth value:
For any $p, q \in S_D$, if $\mathscr{V}(p \Leftrightarrow q) = 1$, then $\mathscr{V}(p) = \mathscr{V}(q)$.
Proof. Trading p for q in Theorem 6 yields $\mathscr{V}(q) \leqslant \mathscr{V}(p)$ for the case $\mathscr{V}(q \Rightarrow p) = 1$. This, together with Theorem 6, entails the desired result.
Remark This corollary is not trivial because it is not restricted to formally true biconditionals.

THEOREM 8.7 For any $p, q \in S_D$, if $\mathscr{V}(p \Rightarrow q) = 1$, then

(i) $\mathscr{V}(p \,\&\, q) = \mathscr{V}(p)$
(ii) $\mathscr{V}(p \vee q) = \mathscr{V}(q)$

Proof. By logic,

$$q \Leftrightarrow q \,\&\, (\neg p \vee p) \Leftrightarrow (\neg p \,\&\, q) \vee (p \,\&\, q)$$
$$q \Leftrightarrow q \vee (\neg p \,\&\, p) \Leftrightarrow (\neg p \vee q) \,\&\, (p \vee q).$$

Taking $\neg p \,\&\, q$ and $p \,\&\, q$ as the variables in Definition 1 leads to (i). Similarly for (ii).

In the case of Theorem 7 the truth value of each statement depends on the degree of truth of the other: it is a case of *alethic dependence*. Alethic dependence subsumes *logical* dependence, which obtains when one of the statements entails the other. Let us now tackle the problem of alethic independence. To this end we shall introduce

DEFINITION 8.5 Let $p, q \in S_D$ with $\mathscr{V}(p) \neq 0$. Then the truth value of q relative to p is defined as

$$\mathscr{V}\left(\frac{q}{p}\right) = \frac{\mathscr{V}(p \,\&\, q)}{\mathscr{V}(p)}$$

(Caution: These are not the conditional truth values characterized in Sec. 2.3. In our view all truth values are conditional, i.e., presuppose some base line or other.)

If $\mathscr{V}(p \Rightarrow q) = 1$, Theorem 7 (i) and Definition 5 entail $\mathscr{V}\left(\frac{q}{p}\right) = 1$. In all other cases $\mathscr{V}\left(\frac{q}{p}\right) \neq 1$. Hence $\mathscr{V}\left(\frac{q}{p}\right) - \mathscr{V}(q)$ is a measure of the strength of alethic dependence. This suggests adopting the following

DEFINITION 8.6 Let $p, q \in S_D$ with $\mathscr{V}(q) \neq 0$. Then

(i) p is *alethically independent* of $q =_{df} \mathscr{V}\left(\frac{q}{p}\right) = \mathscr{V}(p)$.
(ii) p is *alethically dependent* upon q iff p is not alethically independent of q.

This relation of alethic independence is not symmetric but, whenever p is independent of q or conversely, the value of their conjunction is the same, namely the product of their degrees of truth. More explicitly, we have

THEOREM 8.8 If p and q are alethically independent statements in S_D, then

$$\mathscr{V}(p \,\&\, q)=\mathscr{V}(p)\cdot\mathscr{V}(q)$$

Proof. By Definitions 5 and 6.

COROLLARY 8.4 If p and q are alethically independent statements in S_D, then

$$\mathscr{V}(p \vee q)=\mathscr{V}(p)+\mathscr{V}(q)-\mathscr{V}(p)\cdot\mathscr{V}(q).$$

We now have all we need in practice: If p and q are alethically dependent we apply Theorem 8, otherwise we use Theorem 9. For purposes of reference we collect these results in the following chart:

p implies q	*p and q are alethically independent*
$\mathscr{V}(p \,\&\, q)=\mathscr{V}(p)$	$\mathscr{V}(p \,\&\, q)=\mathscr{V}(p)\cdot\mathscr{V}(q)$
$\mathscr{V}(p \vee q)=\mathscr{V}(q)$	$\mathscr{V}(p \vee q)=\mathscr{V}(p)+\mathscr{V}(q)-\mathscr{V}(p)\cdot\mathscr{V}(q).$

An obvious application of these results is the following

THEOREM 8.9 Let T be a scientific theory with n independent assumptions A_i. Then

(i) the degree of truth of the axiom basis equals the product of the partial degrees of truth:

$$\mathscr{V}\left(\bigwedge_{i=1}^{n} A_i\right)=\prod_{i=1}^{n}\mathscr{V}(A_i);$$

(ii) the degree of truth of an assumption conjoined with any of its logical consequences equals the former:

$$\text{If} \quad A_i \vdash t \quad \text{then} \quad \mathscr{V}(A_i \,\&\, t)=\mathscr{V}(A_i).$$

Proof. Part (i) follows from an obvious generalization of Theorem 8 to a conjunction of an arbitrary finite number of independent statements. Part (ii) is an application of Theorem 7 (i).

Since a well organized theory is constituted by a bunch of assumptions and all of their consequences, the preceding theorem justifies adopting the following convention concerning the degree of truth of a scientific theory.

DEFINITION 8.7 The *degree of truth of a scientific theory* equals the product of the truth values of its initial assumptions provided the latter are mutually independent.

This definition elucidates the notion of degree of truth of a theory without however allowing us to compute the degree of truth of any nontrivial theory in factual science. Such a numerical value must go uncomputed. All we can do is to estimate the truth value of a few logical consequences of the axioms conjoined with subsidiary assumptions and empirical data, and see whether the results confirm or infirm the premises, both theoretical and extratheoretical. In fact consider the following rather typical process. (For details see Bunge, 1967a, Ch. 15, and Bunge, 1973b, Ch. 10.)

Initial theoretical assumptions: A_1, A_2.
Additional premises: subsidiary hypothesis s and datum e.
Deduce a testable consequence t*:* $A_1, A_2, s, e \vdash t$.
Produce a new empirical datum e' relevant to t.
Contrast t *with* e'.
Estimate the truth value of t assuming e' to be true.
Infer whether the preceding step confirms or infirms the premises.

The attempt to go upstream and calculate the truth value of the initial assumptions on the basis of the degrees of truth of a few of their consequences conjoined with alien extra assumptions (such as s and e) is chimerical.

We close this subsection by defining some related concepts.

DEFINITION 8.8 Let $T = \langle S, \vdash \rangle$ and $T' = \langle S', \vdash \rangle$ be two theories with a common meaning core, i.e. such that $\mathcal{S}(T) \cap \mathcal{S}(T') \neq \emptyset$ and $\mathcal{R}(T) \cap \cap \mathcal{R}(T') \neq \emptyset$. Then T is *truer* than T' iff there is a 1:1 transformation f of S into S' such that, for every p in T, $\mathcal{V}(p) \geqslant \mathcal{V}(f(p))$.

DEFINITION 8.9 Let $T = \langle S, \vdash \rangle$ and $T' = \langle S', \vdash \rangle$ have a common meaning core. Then T and T' are *alethically equivalent* iff there is a truth preserving transformation $f: S \to S'$, i.e. one that, for every p in T, $\mathcal{V}(p) = = \mathcal{V}(f(p))$.

DEFINITION 8.10 Let T and T' be two alethically equivalent theories. Then T and T' are *semantically equivalent* iff they have the same sense and the same referents.

3.5. *Scientific Inference*

Let us now check whether our theory squares with usual procedures of scientific inference. To begin with the theory condones the scientist's common belief that factual propositions are more or less true rather than either wholly true or totally false. Moreover our theory gives quantitative precision to the idea of partial truth. In particular we have the following code.

Scientific vernacular	*Metascientific jargon*
p is *true*	$\mathcal{V}(p) \doteq 1$
p is *approximately true*	$0 \lll \mathcal{V}(p) < 1$
p is *true to within* $\varepsilon > 0$	$\mathcal{V}(p) = 1 - \varepsilon$
p is *partially true*	$\frac{1}{2} < \mathcal{V}(p) < 1$
p is *false to within* $\varepsilon > 0$	$\mathcal{V}(p) = \varepsilon$
p is *nearly false*	$0 < \mathcal{V}(p) \lll 1$
p is *false*	$\mathcal{V}(p) \doteq 0$
p is *truer than* q	$\mathcal{V}(p) > \mathcal{V}(q)$
p and q *agree to within* $\varepsilon > 0$	$\lvert\mathcal{V}(p) - \mathcal{V}(q)\rvert \leqslant \varepsilon$
p and q *disagree to within* $\varepsilon > 0$	$\lvert\mathcal{V}(p) - \mathcal{V}(q)\rvert \geqslant \varepsilon$,

where '$\doteq$' is a standard symbol for approximate equality, and ε the discrepancy introduced in Sec. 2.3.

Secondly, the theory contains the modus ponens and the modus tollens, which are the cornerstones of the theory and practice of deduction. In fact, if $\mathcal{V}(p \Rightarrow q) = 1$ and $\mathcal{V}(p) = 1$, then by Theorem 7(i) $\mathcal{V}(p \,\&\, q) = \mathcal{V}(p)$. Replacing these values in Definition 1 we get $\mathcal{V}(q) = 1$. Similarly for the modus tollens: setting $\mathcal{V}(q) = 0$ in Theorem 7(ii) yields $\mathcal{V}(p \vee q) = 0$, which substituted into Definition 1 entails $\mathcal{V}(p) = 0$. Similar results are obtained by substituting $\doteq$ for $=$, i.e. for the weakened inference patterns.

Suppose now a conditional is asserted tentatively and its consequent is tested for and found to be true (or false) to within a certain discrepancy ε, where $0 < \varepsilon \lll 1$. That is, we have

(i) *Confirmation*

Assumptions: $\mathscr{V}(p \Rightarrow q) = 1,\ \mathscr{V}(q) = 1 - \varepsilon$
Consequence: $\mathscr{V}(p) = 1 - \varepsilon - \mathscr{V}(\neg p \ \&\ q) \leqslant 1 - \varepsilon$

(ii) *Refutation*
Assumptions: $\mathscr{V}(p \Rightarrow q) = 1,\ \mathscr{V}(q) = \varepsilon$
Consequence: $\mathscr{V}(p) = \varepsilon - \mathscr{V}(\neg p \ \&\ q) \leqslant \varepsilon.$

In words: If q is confirmed, then p can be assigned an *upper bound* less than or at best equal to the value of q. And if q is refuted so is p. In short we retrieve what we knew all along, namely that, whereas confirmation is inconclusive, refutation is pretty clear cut – that is, if we restrict our considerations to pairs of statements.

The previous results hold only for *isolated* statements – the normal situation in inductive logic but an extremely artificial one in science. In live science hypotheses are assigned truth values in the light of both further hypotheses and whole bodies of empirical evidence; likewise the latter are estimated in the light of both further evidence and hypotheses –actually whole theories. When this circumstance is taken into account it becomes possible to obtain bounds other than those calculated above: i.e., confirmation may then be strengthened and refutation weakened – or conversely. In short, the whole body of relevant knowledge is summoned to pass judgment on every single hypothesis and every single datum. It is not that *das Wahre ist dans Ganze* (Hegel), but that the *recognition* of truth and of falsity calls for a whole battery of tests (Bunge, 1961c, 1967a, Vol. II, Ch. 15). The whole may be fully meaningful but, if factual, cannot be totally true.

Finally we note that our theory of truth is contiguous with the theory of scientific inference and, in particular, with the calculus of errors of observation. The latter is in charge of assigning the error or discrepancy ε a numerical value.

3.6. *Comments*

(i) The truth measure $\mathscr{V}$ is a continuous function. But this does not allow one to replace the tautology/nontautology dichotomy by a richer gamut of degrees of *logical* truth and falsity. This Aristotelian dichotomy is very basic: it is built into the very algebra of statements, which is a Boolean algebra with just two selected elements, □ and ▯. *This* is what we take "two valued" to mean in reference to ordinary logic – not that truth values other than 0 and 1 are disallowed. In order to introduce (in a purely syntactic fashion) degrees of analyticity intermediate between tautology and contradiction one would have to modify that algebraic structure – e.g., by increasing the number of distinguished elements of the statement set and characterizing them appropriately. However, it is doubtful that such a reform, though algebraically feasible, would be of interest to the logic of factual science. In any case the truth free formulation of logic we have adopted reminds one that logical truth and falsity, unlike factual truth and falsity, are structural or algebraic – something that the model theoretic presentation of logic tends to becloud. Logics are not truth calculi but entailment calculi.

(ii) The above mentioned continuity of the valuation function $\mathscr{V}$ allows one to consider arbitrarily close approximations to either complete truth or total falsity. A typical example is afforded by any series expansion. When expanding a function as a convergent series and adding up only the first n terms, one makes the error $R_n = |S - S_n|$, where S is the exact but perhaps unknown sum of the series while S_n is the known or knowable sum of its first n terms. As more terms are added the remainder R_n decreases and the truth value of the approximation increases accordingly. Indeed, one may set $\mathscr{V}(S_n \mid S) = |1 - R_n/S|$.

(iii) Our theory enables one to elucidate the intuitive notion of the "asymptotic approaching-to total truth" – sometimes also expressed, misleadingly enough, as "the gradual advance from relative truth to absolute truth". One way of exactifying this notion is, following Reichenbach (1949), to apply the standard notion of convergence of a sequence to a sequence $\langle p_n \mid n \in N \rangle$ of propositions all with the same form and referent – e.g., the successive outcomes of measurements of the electron charge. This can easily be done but is of no great help because we never have infinite sequences of factual statements of that kind, every one of which has been assigned a truth value. More often than not we switch our interest from one family of propositions to another before we have had a chance to form a sequence long enough to suggest (rather than exhibit) any convergence properties (Bunge, 1963a). Moreover, the corresponding sequence of truth values, i.e. $\langle \mathscr{V}(p_n) \mid n \in N \rangle$, could not possibly exhibit convergence in the strict mathematical sense because the terms of the sequence obey no law. Surely there are sets of theories with the same referent and constituting finite increasing sequences, but these are not very long and there is no reason to believe that any of them could continue indefinitely. In any case there is no confirmed theory of knowledge containing a law statement – not just a Panglossian opinion – to the effect that every sequence of hypotheses (or of theories) concerning any given factual referent should converge to total truth. What we do have is a few generalizations with a restricted scope, such as this one: "The truth value of a statistical estimate approaches unity as the sample size approaches the total population". But this approaching of partial to complete truth is not uniform or lawful, hence it cannot be described with the help of the mathematical concept of limit. Thus if the population consists of A's and B's in the same proportion, sampling may result in picking sheer A's during the first half of the time and only B's the rest of the time, so that during the first period the relative frequency of A's will be 1 instead of the long run frequency $\frac{1}{2}$.

(iv) The valuation function $\mathscr{V}$ is "external" to the algebra of propositions. This tallies with the fact that factual truth values are assigned rather than extracted from the propositions themselves by sheer force of analysis. The latter procedure works only, as noted by Leibniz, for logical truths and falsities. The externality of $\mathscr{V}$ to the algebra has the advantage that we might try functions other than the one determined by Definition 1 while keeping the logic intact. Moreover one might study the whole lot of continuous measures on S_D.

(v) Our claim that truth values, if factual, do not inhere in propositions but are bestowed upon them *ab extrinseco* agrees with our treating truth values, when factual, as values of a certain function $\mathscr{V}$ rather than as elements of the algebra of statements. That factual truth values are assigned (and reassigned) rather than disclosed by logical analysis does not entail that our theory elucidates a pragmatic or methodological concept of truth. It does entail that any application of our alethics calls for methodology. A pragmatic concept of truth would be characterized by (*a*) some function $\mathscr{V}_p$, other than our $\mathscr{V}$, on the set $S_D \times P$ of statement-person pairs, and (*b*) certain assumptions concerning the P's, in particular their truth value assignment habits, or else norms.

(vi) We have defined $\mathscr{V}$ on a proper subset S_D of the totality S of statements. That is, we abstain from assigning truth values to a number of statements – e.g. for lack of relevant evidence or for the absence of a proof. (This aspect of our semantics might be approved of by the mathematical intuitionist.) The reason is simple: as a matter of fact most statements in factual science go unvalued. In short, our alethics makes room for what have been called *truth value gaps* – without requiring any change in logic. It only requires the notion of a partial function, which can be construed as a total function on a domain enriched with a fictitious element embodying the undefined (Scott and Strachey, 1971).

(vii) An alternative solution to the problem of the truth value gaps is of course the adoption of intuitionist logic. But this would require the reconstruction of the whole of mathematics in intuitionist terms, since factual science utilizes, in principle, the totality of mathematics. Our solution to the problem is far less expensive.

(viii) Another solution to the problem of the truth value gaps is to adopt some system of three valued logic involving a third truth value – call it "indeterminate". This move was actually proposed in relation to quantum mechanics, which – like every other theory for that matter – contains empirically untestable statements, such as those referring to the "interphenomena" or events assumed to occur between observations (Reichenbach, 1944). But this proposal, like similar proposals made by a number of authors for different nonreasons, is not workable. First, one should be able to put his finger on at least *one* proof in quantum mechanics requiring rules of inference other than those consecrated by ordinary logic. Second, quantum mechanics should then be reaxiomatized on the basis of the alternative logical calculus. None of these two conditions has been satisfied: this logical revolution is still, after four decades, in the initial proclamation stage. And if it became victorious if would asphyxiate quantum mechanics by isolating it from the rest of physics, which would presumably keep its classical logic. In fact, theories with different logics cannot be conjoined, as they must be if they are to become applicable and testable. (See Bunge, 1973a, and 1973b.)

(ix) The metric Boolean algebra introduced by Definition 1 should not be mistaken for a probability measure. For one thing the arguments of $\mathcal{V}$ are individuals (propositions) not sets. (In other words the domain of $\mathcal{V}$, unlike that of Pr, is not a field of sets.) For another, our normalization condition is $\mathcal{V}(p)=1$ for $p \in \square$, not $\mathcal{V}(S_D)=1$ as it should in order to be a probability measure. Thirdly, our theory contains no semantic assumptions asserting that "$\mathcal{V}(p)$" represents the probability of the proposition p – whatever this expression may mean if indeed it means anything. (It cannot mean "the probability of being true", for this would drag us in a circle: the statement "$Pr(\mathcal{V}(p)=1)=r$" involves the concept of total truth.) This semantic trait of our theory should suffice to distinguish it from the various probability (or improbability) theories of truth even if we were to choose exactly the same mathematical formalism.

There is more: whereas our axioms can be applied right away to situations of interest in real science, the axioms of measure theory yield no probabilistic statements without further ado. Indeed, probability theory proper, as distinct from measure theory, starts where the latter leaves off, namely by specifying (constructing) a probability space or space of "events". (Our basic set $[S]$, on the other hand, is sufficiently characterized by saying that it forms a Boolean algebra.) To put it another way: measure theory provides only the foundations of the probability calculus. These foundations cannot be activated to yield probabilistic results, such as the laws of large numbers, unless enriched with some definite *model* of a possible (though idealized) situation, such as the coin model, an urn model, or a Markov chain model. No such model no probability theory proper. (See, e.g., Kolmogoroff, 1963; Feller, 1968.) And such specifically probabilistic devices are as alien to our theory of truth as they are to measure theory: both truth and measure are independent of randomness. Moreover, all such probabilistic models do is to allow one to assign probabilities to the elementary "events" (the individual sets), never to the corresponding statements. The probability statements themselves, if theoretical, are supposed to be totally true and derived in accordance with ordinary logic. (Their factual truth value is of course another matter.) The widespread view that the theory of probability is a generalization of logic involving probability implications and condoning inductive reasoning ignores the perfectly classical logical structure of the theory. There is somewhat more to all this: we shall look into it in the next section.

(x) Our theory of truth explains what is wrong with Frege's conception of a predicate as a function from individuals in a domain D to truth values (Ch. 1, Sec. 1.3). If we simplify, pretending (with Frege) that every proposition can be assigned a truth value (i.e. if we set $S_D = S$) and keep only the two extremes 0 and 1 of the unit interval, we get the functions

$$P: D \to S \quad \text{and} \quad \mathcal{V}: S \to \{0, 1\}.$$

The composition of these functions yield what may be called the *Frege predicate* F corresponding to the predicate P:

$$F =_{df} \mathcal{V} \circ P: D \to \{0, 1\}.$$

What Frege did was to skip the preceding analysis of F into $\mathcal{V}$ and P.

$$\begin{tikzcd}
D \arrow[r, "P"] \arrow[dr, "F"'] & S \arrow[d, "\mathcal{V}"] \\
& \{0,1\}
\end{tikzcd}$$

4. TRUTH ET ALIA

4.1. *Truth and Probability*

Skeptics and empiricists have rightly insisted on the "probable" nature of factual knowledge. But, until recently, they employed the term 'probable' in a nontechnical acceptation, namely as a synonym for either 'uncertain' or 'corrigible' or both. The probability theories of truth proposed over the past six decades appeared to formalize that view by taking the popular 'probable' in the technical sense that the theory of probability assigns to it. More precisely, these theories equate the degree of truth of a factual statement with its probability or with some increasing (or decreasing) function of the latter. They are thus exact philosophical theories. But they are also hollow, for there exists no procedure, other than arbitrary decree, for assigning probabilities to statements. The passion for exactness is certainly a noble passion but, like any other, it may make fools of us.

Both in applied probability theory and in theoretical science the way to assign primary probabilities is to set up some stochastic model of the *factual* system concerned, such as an urn model. (Recall point (ix) in Sec. 3.6.) This procedure does not work for statements, *e.g.*, scientific hypotheses, if only because these are not picked at random, say by extracting them from a hat full of white (true) and black (false) statements. The very concept of randomness, without which the probability theory finds no application, makes no sense for a unique and carefully contrived object such as a scientific hypothesis.

What can often be assigned probabilities are the *facts* referred to by a (factual) probability statement. Thus we may be able to compute, with the help of a *specific* stochastic theory, or to measure, with the help of a *specific* experimental set up, the probability of a certain *event* belonging to a uniform class of factual items, such as rainfalls in some area. But such probability assignments will be correct or incorrect to some extent that is independent of the objective probability of the given event. For example, the probability of a certain nuclear event e may be exceedingly small, say 10^{-24}, while the degree of truth of this probability assignment may be fairly high, say 0.9. That is, we can have $\mathscr{V}[P(e)=10^{-24}]=0.9$. And we can also have the opposite situation, i.e. a high probability value assignment that is nearly false. In short, probabilities of facts and degrees of truth are mutually independent. Hence there is no way of inferring the degree of truth of a statement from the probability of the fact it concerns, nor conversely. Put in another way: There is no supertheory, concerning both a factual domain and a theory about it, that contains law statements relating facts to our knowledge of them. We do not even know whether such lawful connections between facts (random or otherwise) and our knowledge of them could be found.

In conclusion, since the probability (and the information theoretic) theories of truth are not in a position to assign probabilities to statements, we should stop equating the epistemological 'probable' (='uncertain' or 'corrigible') with the semantical 'partially true'. And it would not hurt looking at the way scientists actually estimate the size of the grain of truth of their theories. They may not use the *word* 'truth' (just as they rarely use the term 'cause') because positivism and conventionalism have given it a bad reputation, but they surely use a (presystematic) *concept* of (factual and partial) truth, as is clear from their relentless search for improved representations of facts.

Besides, the concepts of truth are both *more basic* and *more universal* than the concept of probability. In fact we want to be able to say that a given probability assignment (to a fact) is close to (or far from) the truth. And, while all probability statements have presumably some truth value or other, they are only a proper subset of the totality of statements. In short, there is no surrogate for factual truth, and the theory of truth must preside over both stochastic and nonstochastic factual theories. And alethics as well as applied probability presuppose ordinary logic. Practically every theory does.

4.2. *Truth, Meaning and Confirmation*

In Ch. 7, Sec. 2.2 we stipulated that meaning is a property of constructs and that only those expressions in a language that happen to designate constructs are significant. And in this chapter we have agreed that among constructs only statements may, but need not, be assigned a truth value. Figure 8.7 displays these views.

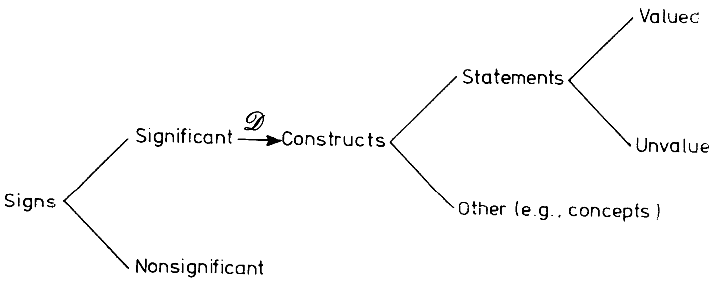


Fig. 8.7. Truth rides on meaning.

Truth depends on meaning (sense together with reference) but not conversely: one and the same factual proposition may be assigned now one truth value, now another, without any change in meaning. (Changes in meaning involve changes in the constructs themselves.) For example, Galilei's refutation of Aristotle's laws of motion did not change the latter's meaning. Galilei could not have accepted either Frege's view that truth values determine meanings or the verifiability doctrine of meaning, according to which meanings are secreted by verification procedures.

However, these doctrines asserting the primacy of truth over meaning survive in various watered down forms. One of these is the view that the truth conditions of a language determine the semantics of the latter. (For the notion of truth condition see Sec. 2.4.) In pragmatic terms: "To give the semantics of a language boils down to stipulating its truth conditions." Let us not quibble over the bizarre belief that languages, rather than theories, come with truth conditions. (For a defense of the alethic neutrality of languages see Ch. 1, Sec. 1.1.) The thesis may fare well in elementary logic but it certainly fails in factual science. First, because in the latter there are hardly any necessary and sufficient conditions under which a statement may be pronounced totally, or even partially, and definitively, or even provisionally, true or false. (Recall Sec. 2.4.) Second, because although every decent scientific theory has a reasonably definite meaning (sense and reference), determined by its basic assumptions, we may be at a loss as to the conditions under which it may be regarded as (approximately) true. In conclusion, the claim that truth conditions determine meanings is as false as the converse thesis. Meaning and truth are equally basic semantic components of propositions.

Nor is truth to be mistaken for confirmation. The two concepts belong in different categories: truth in semantics, confirmation in methodology. But at least in this case taking the one for the other constitutes an intelligent mistake, for confirmation is necessary, though insufficient, to bestow truth values. Indeed, if a statement has been abundantly confirmed by experience and if it coheres with reasonably well corroborated theories, then it may be assigned a truth value close to unity – until new notice. But, as shown by superstition, empirical confirmation alone is insufficient – and refutation may not help. Furthermore empirical tests are sometimes impossible, despite which we may have reasons for allotting a hypothesis a high degree of truth. Figure 8.8 illustrates a frequent situation in theoretical science. (See Bunge, 1967a, I, 5.6.)

What holds for the qualitative concepts of truth and confirmation holds also for their quantitative explicata: confirmation is but an insecure indicator of truth. Thus the statement ⌜All natural numbers are greater than any preassigned natural number⌝ is patently false but, since it has infinitely many confirmers, it could be assigned a maximal degree of confirmation. In sum the two concepts, truth and confirmation, though related, are distinct. A similar relation holds between the semantic con-

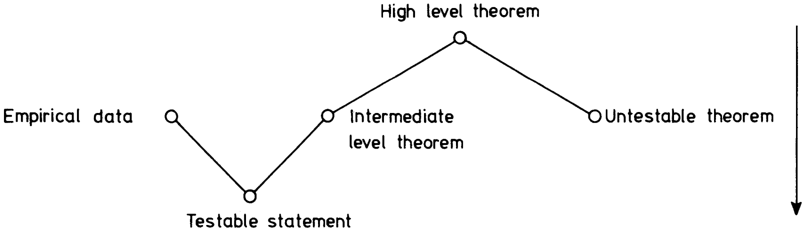


Fig. 8.8. An untestable theorem may be regarded as (more or less) true if it belongs to a theory abounding in corroborated theorems.

cept of truth and the pragmatic (or psychological) concept of acceptance or credence or belief. But this deserves a separate subsection.

4.3. *Truth and Belief*

Except for the dogmatist, truth is not belief: anyone can believe falsities and disbelieve truths – and all of us ignore most truths and most falsities. Let us keep truth and belief distinct – though not necessarily separate. Marry they must if they are to bear fruit or rather edible fruit. The semantic concept of truth and the pragmatic (or psychological) concept of belief (personal or collective) get fused into that of pragmatic truth, or truth for someone. More exactly, there are two such concepts: *personal* (subjective) truth and *collective* (or intersubjective) truth – which, as the fable of the naked king warns us, must not be confused with objective truth. These concepts occur in statements of the form $\ulcorner x$ believes $p\urcorner$, where the substitution instances of x are persons or social groups. The theories dealing with statements of this kind have come to be known as systems of *doxastic logic* and they are closely related to the theories concerning statements of the form $\ulcorner x$ knows that $p\urcorner$, which are the object of the so called *epistemic logic* (Hintikka, 1962).

While there is no doubt that both "logics" are worthy enterprises, it seems clear that they are not *logics* but factual disciplines and consequently in need of empirical tests. Indeed, any belief statement and any knowledge statement of the sort mentioned a while ago concern both propositions and people: they "say" as much or more about persons (or social groups) as about propositions. (They are not metastatements such as $\ulcorner p$ is true$\urcorner$ but object statements with two referents.) This being the case, only psychologists can tell us anything about the truth value of the hypotheses spurned by the doxastic and the epistemic logicians. In other words, although the empirical sciences of belief and knowledge are welcome and moreover long overdue, an *a priori* theory of belief and knowledge stands no chance of being true – hence of contributing to knowledge and compelling our belief in it. To put it another way: Even belief and knowledge statements, i.e., statements of personal (or collective) belief or knowledge, ought to be objective and empirically testable.

Of course objective truths (and falsities) are not Platonic Ideas: *pace* Bolzano there is no such thing as a *Wahrheit an sich*, independent of thinking beings (Bolzano, 1837, I, Sec. 25). Statements in themselves are just useful (as opposed to either idle or evil) fictions. A proposition, viewed from a metaphysical angle, is not an autonomous object (an entity) but perhaps a certain equivalence class of brain processes (thoughts) of some kind. (Note the vagueness, unavoidable at the present stage of cognitive psychology.) And a propositional attitude, such as knowing, believing or doubting a proposition, is another equivalence class of brain processes, this time one concerned with further brain processes. Schematically, we have three levels:

(i) *Class of thoughts* of a kind (judgments).
(ii) *Object proposition* = An equivalence class of thoughts of a kind.
(iii) *Propositional attitude* = An equivalence class of thoughts about thoughts.

   (*a*) *p* is *given* – take it or leave it.
   (*b*) *p* is *examined* with a view to finding out its degree of truth, or its relation to other propositions, or its usefulness to some action, etc.
   (*c*) *p* is *assumed* or hypothesized (not necessarily asserted as true).
   (*d*) *p* is *confirmed* (in particular proved) or *infirmed* (in particular refuted).
   (*e*) *p* is *assigned a truth value.*
   (*f*) *p* is *assigned nonsemantic properties:* systemicity, depth, heuristic power, etc.

(*g*) *p* is *adopted or rejected* for a certain purpose, either theoretical or experimental or practical.

(*h*) *p* is *believed* or *disbelieved* or *left in suspense*.

Etc.

To conclude. Truth and belief are heterogeneous categories. Moreover, *pace* the pragmatists (e.g. Zinov'ev, 1973), truth is not definable in terms of acceptance or credence. Rather on the contrary, to anyone but a mystic, believing *p* consists in admitting *p as true*, at least to a considerable degree at the time of thinking *p*. Which leads us to considering the relation of truth to time.

4.4. *Truth and Time*

If factual truth were an intrinsic property of propositions, as is the case with logical truth, then it would be timeless. And even if it were less than this, namely mere mathematical truth, it would be timeless as well: the property of being true in a model is neither eroded nor accrued in the course of time, because models (in the model theoretic acceptation of the term) are themselves timeless objects. Alas, factual statements are more complicated than this: in addition to being true in some model they must be (sufficiently) true of the world, which is largely beyond our control. And they are assigned truth values on the strength of both conceptual and empirical operations of varying degrees of refinement. The upshot is that factual truth values vary in the course of time. In this sense *Veritas filia temporis* (cf. Bunge, 1967a, II, 10.5.)

The time dependence of factual truth value assignments summarizes a complex process of testing for the truth. This process involves flesh and blood scientists handling a changing array of conceptual and physical tools. Although every one of the inputs participating in this process may be lawful, the output, i.e. the trajectory or curve of the successive truth values does not seem to satisfy any law: it is a time series, at most a trend. (For the concepts of trend and law see Bunge, 1967a, I, 6.2 and 6.6.) This does not mean that truth value assignments and reassignments are capricious, hence that there is no such thing as objective truth. All it means is that, as with most historical processes, sequences of truth value assignments are the outcome of the interplay of numerous factors, some of which escape our control.

What about predictions: do they have a truth value before the facts they refer to come to pass? According to Aristotle (*De interpretatione* Book 9) only propositions about actuals are either true or false. On the other hand propositions about contingent futures have no definite truth values. It might be objected that the predictions computed in factual science do have definite truth values since they are entailed by premises taken to be true. But this would be incorrect: we need not, and usually do not, *assert* our hypotheses but just process them to find out what they entail, and wait and see how some of these consequences fare. That is, we may and often must abstain from assigning any truth values to our predictions. So Aristotle would seem to be right.

However, things are not this simple. If the possibles in question are not unique such as tomorrow's naval battle, but get actualized once and again under suitable conditions, then we may subject them to theory. And we may subsequently contrast some statements in the theory with actuality, or the unfolding of potentiality (to continue Aristotle's train of thought), thus being able to evaluate those propositions. This holds in particular, though not exclusively, for stochastic theories: here we contrast calculated probabilities and their subordinates (averages, mean fluctuations, etc.) with observed frequencies and their relatives. But possibles are not the exclusive property of stochastic theories: all scientific theories concern possible things, possible properties, possible states, possible changes of state. (However, science does not need modal logic. It construes "a possible factual item of kind F" as "an arbitrary member of the set F of factual items", and "a probable fact" as a fact with a definite probability. Hence scientific theory, though concerned with possibility, is strictly truth functional.) Only the applications and tests of scientific theories concern actuals. And no sooner are actuals involved, than we can proceed to an effective (yet provisional) determination of truth values. The foregoing discussion is summarized in Figure 8.9.

5. CLOSING REMARKS

A theory of factual truth is often expected to accomplish at least as much as a theory of formal truth, viz.,

(i) offering a tidy definition of "factually true statement";
(ii) laying down universal truth conditions (criteria); and

(iii) giving rules for computing the truth value of any truth functional compound, such as a conjunction, in terms of the truth values of its constituents.

Our theory of truth fails, nay refuses to undertake the first two tasks and performs only the third. In fact we hold (Sec. 4.2) that the concept of factual truth is much too basic and universal to be demoted to the rank of a defined concept. (And in any event all the attempts to eliminate it in favor of alternative concepts, such as those of satisfaction, probability, information, and confirmation, have failed.) The most we can do for the concept of factual truth is

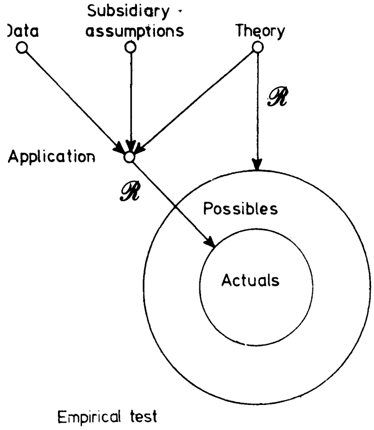


Fig. 8.9. Only the applications and tests of a scientific theory concern actuals: every general theory is about possibles. The outcomes of tests (on actuals) enable one to assign truth values to statements (whether deterministic or stochastic) concerning possibles.

(i) to give an *informal* characterization (Sec. 1);

(ii) to show how it is *used* in scientific practice (Sec. 2); and

(iii) to define it *implicitly* by a set of postulates (Sec. 3).

As to factual truth conditions, we have pointed out that (*a*) in science there are no cut and dried truth conditions in the simple form of equivalences, and (*b*) the complex, regional and changing clues for estimating factual truth values must remain the responsibility of science. Here the semanticist must be a spectator and student not a law giver.

Our theory of factual truth, incomplete though it may look to the philosopher accustomed to the white or black situations encountered in logic, does seem to elucidate in an exact way the concept of degree of truth used by applied mathematicians and scientists. However, we do not claim that it is the best possible (the truest) theory. There may well be alternative measures $\mathscr{V}$ that supply a more adequate explication of the intuitive concept of partial factual truth. (Recall Sec. 3.6, point (iv).) Still, any such alternatives should presuppose classical logic, which is the one inherent in science. And they should satisfy the same desideratum, namely that of supplying an exact version of the imprecise notion of approximate truth (or falsity) of fact.

Note finally that our alethics takes care of the notion of *partial explanation*, which is becoming prominent in the philosophy of science (Scheibe, 1973). A fundamental theory is said to provide only a *partial explanation* of the empirical generalizations (or the phenomenological theory) that motivated the construction of the theory. For example, celestial mechanics does not retrieve Kepler's laws but gives only an "approximate explanation" of them because, strictly speaking, the sun does not stand still and the planets exert perturbations upon one another, all of which complicates the actual planet trajectories far beyond Kepler's simple ellipses. As a matter of fact the explanation of the latter supplied by Newton's celestial mechanics is rigorous: what are not exact, or rather not completely true, are the subsidiary assumptions that the solar mass is infinite and that each planet is acted upon only by the sun. Every application of any theory involves simplifying assumptions like these. Thus in the elementary theory of the simple pendulum one assumes that the oscillation amplitude is small and so he obtains the laws of Galilei and Huyghens, known to be only partially true. The deduction is exact: what is approximate is the simplifying assumption and consequently the conjunction of this subsidiary condition with the general assumptions of the theory. To sum up, if we accept the general concept of partial and relative truth of fact we need not introduce the *ad hoc* concept of approximate explanation. Or at least the latter can be defined in terms of the former.

This closes our exposé of the basic theories in our semantics of science. The rest of the book deals with applications (Ch. 9) and border issues (Ch. 10).

CHAPTER 9

OFFSHOOTS

The ideas on meaning and truth presented and discussed in the preceding chapters may be applied to a variety of problems in philosophical semantics, whether pure (general) or applied (to, e.g., scientific theories). We have dealt with sundry problems in applied semantics throughout the foregoing chapters. Let us now face three problems in pure or general semantics, namely those of extension, vagueness, and definite description: in our system all of them presuppose the theories of reference, sense, and truth.

1. EXTENSION

1.1. *The Problem*

An extension is the extension of some predicate. Thus the extension of "– is a mountain" is the class of mountains. And every extension is a set. But not every set is the extension of some predicate. For example there seems to be no (interesting) predicate corresponding to the set {China, d/dx}. Nor does every construct have an extension: only those constructs can be assigned an extension for which the concept of truth makes sense. It makes sense to ask what are the objects for which a certain predicate P holds: the collection of such individuals is the extension of P, or $\mathscr{E}(P)=\{x \mid Px\}$ if P is a unary predicate. Individual concepts, such as "Archimedes", have a referent but not an extension. (Individual variables, such as "here" or "we", fare even worse, as they have no fixed referents.) Nor do sets have an extension – or, if preferred, they are their own extensions. (Sets are in fact the only purely extensional objects.) Not even closed formulas, whether simple such as $\ulcorner Pa \urcorner$, or complex such as $\ulcorner (x)(\exists y)\, Pxy \urcorner$, are usually assigned an extension – although they might. Normally only predicates are assigned an extension or domain of application. At any rate we shall confine the theory of extensions to predicates.

The concept of extension must be elucidated by a theory of extensions. The theory of reference cannot do this job because the concepts of reference and of extension are quite different and probably not interdefinable. For one thing the notion of reference does not presuppose the concept of truth, which the concept of extension does. For another the extension of an n-ary predicate is a set of ordered n-tuples, whereas the reference class of the same relation is the scrambled set of the individuals concerned. Thus

$$\mathscr{E}(\text{loves}) = \{\langle \text{Abélard, Héloïse} \rangle, \langle \text{Dante, Beatrice} \rangle, \ldots$$
$$\mathscr{R}(\text{loves}) = \{\text{Abélard, Héloïse, Dante}, \ldots\}.$$

Prima facie there are two other theories that could claim to constitute a theory of extensions each, namely ordinary logic and set theory. But the former is truth functional rather than "purely extensional", as we argued in Ch. 4, Sec. 1. Logic determines neither intensions nor extensions: it leaves them indeterminate. Only the semantics of logic introduces extensions when supplying models for the logical calculi. (*A fortiori* logic cannot be construed as universal semantics – *pace* Bar-Hillel (1970).)

As to the claim that set theory is the theory of extensions, its force depends upon the version of the theory one has in mind. Certainly it does not hold for the von Neumann-Bernays-Gödel theory, which does not involve the concept of a predicate, whence it cannot be construed as dealing with the extension of predicates: in that theory a class is an object on its own. (The rationale for this is not that "mathematics has no need for non-classes, like cows or molecules" (Mendelson, 1963, p. 160), but rather that, *once generated*, a set may be handled as an object in its own right. But if the problem is to determine or characterize a particular infinite set, then there is no way except to seize on some property and use the principle of abstraction or its rigorous version, the *Aussonderung* axiom – which is absent from the von Neumann-Bernays-Gödel theory.) As to the Zermelo-Skolem-Fraenkel version of set theory, which does involve the notion of a predicate, it is certainly concerned with extensions since it contains the golden bridge between predicates and sets – namely the *Aussonderung* postulate. However, the central goal of this theory is not to elucidate the notion of extension by relating it to that of truth and distinguishing it from that of reference – which is what semantics is mainly interested in.

We need then a separate theory of extensions, distinct from both logic and set theory though subordinated to them. In fact we need two theories of extension if we are to cope with factual science: (*a*) a theory of *strict* extensions, concerning clear cut (decidable) predicates as well as total (not just approximate) truth, and (*b*) a theory of *lax* extensions, concerning predicates that are either inherently vague or whose instances are known imperfectly. The present section is devoted to strict extensions, the next to lax extensions.

1.2. *Strict Extension: Definition*

While the referents of a predicate are the objects it points to, whether truly or not (Ch. 2), the extension of the predicate is the collection of objects for which it holds and in the order in which it holds. More precisely the extension of a predicate P is the *truth set*, or *solution set*, or *graph* of P. This graph is included in the domain of definition of P. *Example 1* Whereas the reference class of "is literate" is mankind, its extension is the subset of persons who actually read and write. *Example 2* Let P be a binary predicate and, more specifically, one whose relata are tied by P in this way: $Pxy = (f(x) = y)$, for x and y in the real line R. The extension of P is the graph of f – a curve in the plane $R \times R$. See Figure 9.1.

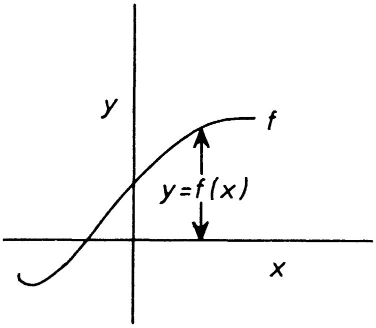


Fig. 9.1. The curve represents the extension of the binary predicate $P = X \times Y \to S$ such that $Pxy = (f(x) = y)$. Indeed, $\mathscr{E}(P) = \{\langle x, y\rangle \in X \times Y \mid y = f(x)\}$.

These ideas are spelled out by the following

DEFINITION 9.1 Let $\mathbb{P}$ be the family of all predicates defined on a domain

$D=A_1 \times A_2 \times \ldots \times A_n$. Then the function

$$\mathscr{E}=\mathbb{P}\to\mathscr{P}(D)$$

such that, for any predicate $P:D\to$Statements, belonging to the predicate family $\mathbb{P}$, $\mathscr{E}$ takes the value

$$\mathscr{E}(P)=\{\langle x_1, x_2, \ldots, x_n\rangle \in A_1 \times A_2 \times \ldots \times A_n \mid Px_1x_2\ldots x_n\},$$

is called the *extension function for P* and its values *extensions* (or *extension values*).

Example 1 The extension of "is a planet of the Sun" is the set of the eight planets certified to be such. The ghostly Pluto is among the referents of the predicate and even in its lax extension but it does not qualify, at the time of writing, as a member of its strict extension. *Example 2* A function f and its restriction $f \mid A$ to a set A, though cointensive, have different extensions: $\mathscr{E}(f \mid A) \subset \mathscr{E}(f)$. Obviously this holds, in particular, when A is a singleton.

Remark 1 The extension function $\mathscr{E}$ maps predicates neither into nor onto their domains: the latter contain regions that correspond to different predicates or to none. *Remark 2* Our very notation shows that we reject the extensionalist identification of a predicate with its graph or extension – for the reasons given in Ch. 4 Sec. 1.2. Thus instead of writing '$\langle x, y\rangle \in R$' for a binary relation R, we prefer to write '$\langle x, y\rangle \in \mathscr{E}(R)$'. This difference, tiny as it is for practical purposes, is important for the foundations and the philosophy of mathematics as well as for semantics. And it avoids funny expressions such as '$= \,=\mathrm{diag} X \times X$' and '$= \subset \subseteq$'. *Remark 3* The distance between extensions and reference classes can be considerably shortened if, instead of conceiving n-tuples as sets of sets (*à la* Wiener and Kuratowski), we construe them as individuals (points). To this end we may adopt the original Bourbaki point of view, now championed by Mac Lane, according to which an ordered pair is a complex individual characterized by the single axiom (not definition): $\langle x, y\rangle = \langle x', y'\rangle$ iff $x=x'$ and $y=y'$. *Remark 4* Definition 1 bridges Platonism ("Only forms are real") and nominalism ("Only individuals are real"). Or rather it shows that the Platonism/nominalism divide is not a reasonable one, as there are neither pure forms nor formless individuals. Every form is the form of something and every individual exemplifies some form or other. *Remark 5* Because truth is relative, so is extension. The extension of a factual predicate grows or shrinks with the advancement of knowledge.

1.3. *Some Consequences*

First an immediate consequence of Definition 1 in Sec. 1.2:

COROLLARY 9.1 The extension of a predicate is included in its domain:

$$\text{If} \quad P{:}D \to \text{Statements}, \quad \text{then} \quad \mathscr{E}(P) \subseteq D.$$

Now the main result. But before stating it we must recall that a compound predicate is defined only on the overlap of its components (Ch. 1, Sec. 1.3).

THEOREM 9.1 The extension of a Boolean function of predicates equals the corresponding Boolean function of their images under the extension function. I.e., if P and Q are predicates defined on a common domain, then

(i) $\mathscr{E}(\neg P) = \overline{\mathscr{E}(P)}$;
(ii) $\mathscr{E}(P \,\&\, Q) = \mathscr{E}(P) \cap \mathscr{E}(Q)$;
(iii) $\mathscr{E}(P \vee Q) = \mathscr{E}(P) \cup \mathscr{E}(Q) \supseteq \mathscr{E}(P \,\&\, Q)$;
(iv) $\mathscr{E}(P \Rightarrow Q) = \overline{\mathscr{E}(P)} \cup \mathscr{E}(Q)$.

Proof. For the sake of simplicity let us confine ourselves to unary predicates defined on a domain D. By Definition 1,

$$\mathscr{E}(\neg P) = \{x \mid \neg Px\} = D - \mathscr{E}(P) = \overline{\mathscr{E}(P)},$$

which proves (i). As to (ii), since $P \,\&\, Q$ is defined on a common domain, it can be treated as a single predicate with values $(P \,\&\, Q)\,x$, where $x \in D$:

$$\mathscr{E}(P \,\&\, Q) = \{x \mid (P \,\&\, Q)\,x\} = \{x \mid Px \,\&\, Qx\} = \\ = \{x \mid Px\} \cap \{y \mid Qy\} = \mathscr{E}(P) \cap \mathscr{E}(Q).$$

Analogously for (iii). Finally, (iv) is proved by replacing $P \Rightarrow Q$ by $\neg P \vee Q$ and using (i) and (ii).

Example 1 $\mathscr{E}(\text{nonliving}) = \overline{\text{Living beings}}$.

Example 2 $\mathscr{E}(\text{small circular}) = \mathscr{E}(\text{small}) \cap \mathscr{E}(\text{circular})$.

On the other hand "is a circular square", which refers to plane figures, has a nil extension. (And it is meaningful: recall Ch. 7, Sec. 2.2.)

Remark 1 The condition of Theorem 1, that the component predicates be defined on a common domain, disqualifies the composition (e.g., conjunction) of referentially heterogeneous predicates such as "metallic" and "jealous". Hence we cannot even say of them that they have empty extensions: they have no extensions at all – nor do they have an intension, as we learned in Ch. 4, Sec. 2.2. *Remark 2* Theorem 1(ii) is the basis of the usual procedure for constructing the (extensionally) smallest of all the predicates satisfying a certain condition – namely to conjoin them. For example, consider the class of equivalence relations $\sim_n$, with $n \in N$, on a given set S. The smallest of them all is $\sim_0 = \bigwedge_n \sim_n$. This is the relation that holds only between an element of S and itself – i.e., strict identity.

COROLLARY 9.2 Double negation restores the original extension:

$$\text{If } P \text{ is a predicate, then } \mathscr{E}(\neg\neg P) = \mathscr{E}(P).$$

COROLLARY 9.3 Inconsistent predicates are extensionally void and tautological ones are universal: If P is a predicate with domain D, then

(i) $\mathscr{E}(P \,\&\, \neg P) = \emptyset$;
(ii) $\mathscr{E}(P \vee \neg P) = D$.

Proof. Let $Q = \neg P$ in Theorem 1 (ii) and (iii) and use (i). To generalize to arbitrary tautologies and contradictions use the theorem that all tautologies are equivalent, hence coextensive.

Remark 1 All of the above formulas concern predicate variables. Being general they hold for any predicate constant regardless of its reference. For example, if P is assigned the value "philosopher" in Corollary 3, then P refers to persons and is true of philosophers, while $P \,\&\, \neg P$ and $P \vee \neg P$ still refer to persons but now $P \,\&\, \neg P$ "applies" to nothing whereas $P \vee \neg P$ is true of everything. *Remark 2* Corollary 3(ii) and its generalization to an arbitrary tautological predicate is a reason for maintaining that logic is universal strictly speaking, i.e., that it "applies" to everything, or that it holds "in all possible worlds". This could be accepted provided it be not interpreted as the assertion that logic is a sort of universal physics (Gonseth, 1938, p. 20) or metaphysics (Scholz, 1969, pp. 399 ff.). This interpretation is mistaken because (*a*) the referents of the above formulas are predicates (universals) not things, and (*b*) the formulas and rules of logic are consistent with mutually incompatible views of the world. It is the business of metaphysics (or ontology), not of logic, to discover the bricks and plans, if any, of the real world (see Ch. 10, Sec. 4). Logic is not metaphysics, it does not depend on the latter, and it suggests no particular metaphysics (Nagel, 1956). More on this in Ch. 10, Sec. 4.2 and in Bunge (1974a).

An equivalent formulation of Theorem 1 is this: Attributes are isomorphic with their extensions. This is why one can think either in terms of attributes or of their extensions, as well as switch back and forth between them. More precisely, Theorem 1 can be reworded thus:

THEOREM 9.2 Let $\mathbb{P}$ be a family of predicates defined on a common domain D. Then the following holds:

(i) The structure $\mathcal{P} = \langle \mathbb{P}, \square, \mathrm{▯}, \vee, \&, \neg \rangle$, where $\square \in \mathbb{P}$ is the null predicate (the one with null extension) and $\mathrm{▯} \in \mathbb{P}$ is the universal predicate (the one that applies to every point in D), is a Boolean algebra.

(ii) The algebra of sets on D, i.e., $\mathcal{D} = \langle \mathcal{P}(D), \emptyset, D, \cup, \cap, ^{-} \rangle$, is a Boolean algebra.

(iii) $\mathcal{P}$ and $\mathcal{D}$ are isomorphic.

This result had to be stated explicitly because occasionally (e.g., Birkhoff, 1961, p. 189) one reads that there is a dual isomorphism (i.e., an anti-isomorphism) between the algebra of attributes and the one of classes. The duality holds between predicates and their extensions: recall Fig. 4.2 in Ch. 4, Sec. 2.3.

Finally a word on the relation between extension and reference. Except in the case of unary predicates this relation is not simple, because (*a*) extension is reference *cum* truth, and (*b*) whereas reference classes are classes of individuals, extensions are sets of n-tuples. A simple relation obtains only between extension and adequate or correct reference. The latter concept is introduced by

DEFINITION 9.2 The *adequate reference* $\mathcal{R}_p^+(P)$ of a predicate P equals the union of the projections of its extension $\mathcal{E}(P)$ on the cartesian factors of the domain of P. In particular, for a binary predicate P defined on $A \times B$,

$$\mathcal{R}_p^+(P) = p_A \mathcal{E}(P) \cup p_B \mathcal{E}(P) \subseteq \mathcal{R}_p(P).$$

1.4. *Comparing Extensions*

Since Definition 1 and Theorem 1 help us to compute individual extensions, we can *a fortiori* compare them. In particular we can ascertain whether two predicates are coextensive in the sense specified by

DEFINITION 9.3 Let P and Q be any predicates. Then P and Q are said to be *extensionally equivalent*, or *coextensive*, just in case they have the same extension:

$$P \sim_e Q =_{df} \mathscr{E}(P) = \mathscr{E}(Q).$$

Remark 1 Identical predicates are coextensive but the converse is false. Thus "lighter than" and "cheaper than" are coextensive in the collection of materials of a given kind but they are intensionally different. If extensionalism were right then coextensives should be identical.
Remark 2 Consequently Leibniz' principle, *"Eadem sunt, quae sibi mutuo substitui possunt, salva veritate"*, is false. In fact let p and p' be two propositions differing only in that the predicate $P' \neq P$ occurs in p' in exactly the same syntactical place P occupies in p. Assume further that P and P' are coextensive. Then p and p' will have the same truth value although different senses: therefore p and p' *non sunt eadem*. Truth values do remain invariant under substitution of coextensives – but truth is not everything and does not determine meaning.

From Definition 3 it is obvious that $\sim_e$ is an equivalence relation. Hence it generates equivalence classes according to

DEFINITION 9.4 Let $\mathbb{P}$ be a family of predicates. Then the class of $\sim_e$-relatives of $P \in \mathbb{P}$ is the subset of $\mathbb{P}$ formed by the equivalence class generated by P, i.e.,

$$[P] = \{Q \in \mathbb{P} \mid Q \sim_e P\}.$$

We can now state a proposition that in a way is the converse of the principle of abstraction or separation of set theory, namely the one predicate-one set axiom. It is the following

PROPOSITION An arbitrary set A determines an equivalence class of predicates under the relation of equiextensionality, namely,

$$\emptyset \subseteq \{Q \in \mathbb{P} \mid \mathscr{E}(Q) = A\} \subseteq \mathbb{P}.$$

Such classes of coextensive predicates need have no structure whatsoever. However the whole lot of such classes does have a definite structure, as will be seen in Sec. 1.5. But we can do better than grouping coextensive predicates: we can compare predicates that are not coextensive provided they are of the same rank. Thus "ellipse" is extensionally included in "conic section", which is in turn extensionally included in "plane figure". In general we have

DEFINITION 9.5 Let P and Q be n-ary predicates. Then P is *extensionally included in* Q iff the graph of P is included in the graph of Q:

$$P \preccurlyeq_e Q =_{df} \mathscr{E}(P) \subseteq \mathscr{E}(Q).$$

We are now in a position to state and prove

THEOREM 9.3 If P and Q are n-ary predicates such that $P \Rightarrow Q$, then P is extensionally included in Q and conversely:

$$P \Rightarrow Q \quad \text{iff} \quad P \preccurlyeq_e Q, \quad \text{i.e.,} \quad \mathscr{E}(P) \subseteq \mathscr{E}(Q).$$

Proof. Assume that Pa holds. Hence $a \in \mathscr{E}(P)$. By hypothesis, if Pa then Qa. Since we did assume Pa, it follows that Qa, which amounts to $a \in \mathscr{E}(Q)$. But a is an arbitrary solution of Px, hence $\mathscr{E}(P) \subseteq \mathscr{E}(Q)$. The converse is proved similarly.

COROLLARY 9.4 Equivalents are coextensive:

$$P \Leftrightarrow Q \text{ iff } P \sim_e Q, \quad \text{i.e.,} \quad \mathscr{E}(P) = \mathscr{E}(Q).$$

THEOREM 9.4 If P entails Q then the extension of P is contained in the extension of Q, i.e., if $P \Rightarrow Q$ is a tautologous predicate then $\mathscr{E}(P) \subseteq \mathscr{E}(Q)$.

Proof. By Theorem 1(iv), $\mathscr{E}(P \Rightarrow Q) = \overline{\mathscr{E}(P)} \cup \mathscr{E}(Q)$ and, by Definition 1, when $P \Rightarrow Q$ is tautologous $\mathscr{E}(P \Rightarrow Q) = D$. Hence $\overline{\mathscr{E}(P)} \cup \mathscr{E}(Q) = D$, which amounts to $\mathscr{E}(P) \subseteq \mathscr{E}(Q)$.

Let us finally consider the extreme cases of null and maximal extension. More explicitly, we introduce the following two concepts:

DEFINITION 9.6 (i) A predicate with null extension is called *minimal.* (ii) A predicate whose extension coincides with its domain of definition D is said to be *maximal relative to D.*

Example 1 The predicate P such that $Px =_{df} (x^2 = -1)$ & x is a real number, is minimal, for it is satisfied by no real number: $\mathscr{E}(P) = \emptyset$. *Example 2* The predicate P such that $Px =_{df} ((x-1)(x+1) = x^2 - 1)$, is maximal in the field of complex numbers, as it holds for every x in it: $\mathscr{E}(P) = \mathbb{C}$.

THEOREM 9.5 (i) There are infinitely many extensionally minimal predicates of a given rank. (ii) A minimal predicate implies any predicate of the same rank. (iii) There are infinitely many extensionally maximal predicates of a given rank. (iv) A maximal predicate is implied by any predicate of the same rank.

Proof. The first part follows from Corollary 3(i) and the third from Corollary 3(ii). The other two parts follow from Theorem 3 and Definition 6 by recalling that the empty set is included in any set.

In logic and mathematics a minimal predicate is an impossible one: not so in factual science. Here we find plenty of minimal predicates, such as "perfectly rigid", with nonempty reference classes as well as nonvoid senses.

1.5. *Algebraic Matters*

Consider a nonempty set $\mathbb{P}$ of predicates defined on a common domain D. Moreover assume that $\mathbb{P}$ is closed under negation, conjunction, and disjunction. The relation $\sim_e$ of coextension, introduced by Definition 3, effects a partition of the family $\mathbb{P}$ into mutually disjoint equivalence classes $[P]$ characterized by Definition 4. Call

$$\mathbb{P}/\sim_e = \{[P] \mid P \in \mathbb{P}\}$$

the family of all the classes of coextensives in $\mathbb{P}$. We define a partial order $\sqsubseteq$ in this quotient set with the help of the relation $\preccurlyeq_e$ of extensional inclusion introduced by Definition 5:

DEFINITION 9.7 If $P, Q \in \mathbb{P}$, then

$$[P] \sqsubseteq [Q] \quad \text{iff} \quad P \preccurlyeq_e Q.$$

Since $\mathbb{P}$ is partially ordered by $\sqsubseteq$, $\langle \mathbb{P}/\sim_e, \sqsubseteq \rangle$ is a poset itself. Moreover we turn it into a richer structure, namely a lattice, upon defining the lattice operations $\sqcap$ and $\sqcup$ in the following way:

$$[P] \sqcap [Q] =_{df} [P \,\&\, Q], \qquad [P] \sqcup [Q] =_{df} [P \vee Q].$$

In fact consider the class $[P \,\&\, Q]$ of all the predicates coextensive with $P \,\&\, Q$. Since $P \,\&\, Q$ entails P, $[P \,\&\, Q] \sqsubseteq [P]$ by Theorem 3 and Definition 7. Exchanging P for Q we also obtain $[P \,\&\, Q] \sqsubseteq [Q]$. This proves that $[P \,\&\, Q]$ is a lower bound of the subset $\{[P], [Q]\}$ of $\mathbb{P}/\sim_e$. Furthermore $[P \,\&\, Q]$ is the greatest lower bound, or infimum, of that subset. In fact let $[R]$ be a lower bound of the latter, i.e., assume that $[R] \sqsubseteq [P]$ and $[R] \sqsubseteq [Q]$. Then, again by Theorem 3 and Definition 7, $R \Rightarrow P$ and $R \Rightarrow Q$. Now, by logic $R \Rightarrow P \,\&\, Q$, whence $[R] \sqsubseteq [P \,\&\, Q]$. That is, $[P \,\&\, Q] = [P] \sqcap [Q]$ is, indeed, the infimum of $\{[P], [Q]\}$. We proceed analogously with $[P \vee Q]$. This time we start from the logical axiom $\ulcorner P \Rightarrow P \vee Q \urcorner$, to derive $[P] \sqsubseteq [P \vee Q]$ and $[Q] \sqsubseteq [P \vee Q]$. which proves $[P \vee Q]$ to be an upper bound of $\{[P], [Q]\}$. It can also be shown that $[P \vee Q]$ is the least upper bound, i.e., the supremum of the subset of coextensives concerned. Furthermore it can be proved that the equivalence classes in $\mathbb{P}/\sim_e$ inherit the distributivity characteristic of the predicates themselves. Since all of this is proved for arbitrary elements of $\mathbb{P}$ we conclude that the family of classes of coextensives forms a distributive lattice. More precisely, we have

THEOREM 9.6 Let $\mathbb{P}$ be a nonempty set of predicates defined on a common domain and let $\mathbb{P}$ be closed under negation, conjunction and disjunction. Furthermore call $\sim_e$ the relation of equal extension. Then the structure $\langle \mathbb{P}/\sim_e, \sqcap, \sqcup \rangle$ is a distributive lattice.

Moreover this lattice contains a null element $\square$ and a unit element ▯, i.e., equivalence classes such that, for every class $[P]$ of coextensives,

$$\begin{aligned} &[P] \sqcap \square = \square \sqcap [P] = \square, \qquad &&[P] \sqcup \square = \square \sqcup [P] = [P] \\ &[P] \sqcap \text{▯} = \text{▯} \sqcap [P] = [P], \qquad &&[P] \sqcup \text{▯} = \text{▯} \sqcup [P] = \text{▯}. \end{aligned}$$

Indeed, since $\mathbb{P}$ is closed under negation and disjunction, we shall find in it at least one tautologous predicate T – one of the infinitely many maximal predicates discussed towards the end of the previous subsection. Then any other predicate S in the family will entail T, i.e., $\vdash S \Rightarrow T$.

Consequently $[S]\sqsubseteq[T]$. This proves $[T]$ to be the maximal element of $\mathbb{P}/\sim_e$: call it ▯. Similarly for a contradictory predicate $\neg T$: for every predicate S in $\mathbb{P}$, $\vdash\neg T\Rightarrow S$. Hence $[\neg T]\sqsubseteq[S]$. This proves that $[\neg T]$ is the minimal element of $\mathbb{P}/\sim_e$: call it $\square$. Finally we can build the complement $\overline{[P]}=[\neg P]$ of any class $[P]$ of coextensives: it is such that

$$[P]\sqcap\overline{[P]}=\square,\qquad [P]\sqcup\overline{[P]}=\text{▯}.$$

In short, we have proved that our distributive lattice is complemented and comes with a universal and a null element – in sum, that it is a Boolean algebra. To put it explicitly:

THEOREM 9.7 Let $\mathbb{P}/\sim_e$ be a family of coextensive predicates. Then the structure $\langle\mathbb{P}/\sim_e, \square, \text{▯}, \sqcap, \sqcup, {}^-\rangle$ is a Boolean algebra.

This particular Boolean algebra may be called the *Lindenbaum algebra of predicates*, by analogy with the Lindenbaum algebra of propositions. If we disregard all differences between individual predicates and propositions except their differences in extension, we shall limit our attention to their respective Lindenbaum algebras. The extensionalist thesis is, in a nutshell, that only such algebras matter. Our view is, on the other hand, that extensions constitute just one aspect, and not even a basic one, of the concepts of the predicate type. Semantics must investigate all the sides and must show how they are related. We proceed to exhibit the relation between extension and intension.

1.6. *Extension and Intension: The Inverse Law*

The richer a concept the smaller its coverage. Thus since the concept of solid is richer than that of body, the set of solids should be included in the class of bodies. The "inverse law" can be stated in set theoretic terms (Bunge, 1967a, I, p. 68) and can now be proved with the help of the calculus of intensions in Ch. 4 and the concept of extension studied in the previous subsections. It consists in the following

THEOREM 9.8 For any two predicates P and Q of the same rank (hence extensionally comparable)

(ii) If $\mathcal{I}(P) \subset \mathcal{I}(Q)$ then $\mathcal{E}(P) \supseteq \mathcal{E}(Q)$.

Proof of (i). Suppose the consequent of (i) to be false. Then we can "balance" the "inequation" by setting $P = Q$ & R, where R is a nontautologous predicate such that $\mathcal{E}(P) = \mathcal{E}(Q \,\&\, R)$. By our calculus of intensions (Def. 1(i), in Ch. 4, Sec. 2.2), $\mathcal{I}(P) = \mathcal{I}(Q) \cup \mathcal{I}(R) \supseteq \mathcal{I}(Q)$, contrary to hypothesis.

Proof of (ii). Assume $\mathcal{I}(P) \subset \mathcal{I}(Q)$. Then $\mathcal{I}(Q) = \mathcal{I}(P) \cup X$, with X a nonempty set. Since by hypothesis $\mathcal{I}$ is onto, there exists a third predicate R such that $\mathcal{I}(R) = X$. That is, $\mathcal{I}(Q) = \mathcal{I}(P) \cup \mathcal{I}(R)$. And, again by our calculus of intensions, $\mathcal{I}(Q) = \mathcal{I}(P \,\&\, R)$. Now, the extension of P & R is, by the calculus of extensions, $\mathcal{E}(P \,\&\, R) = \mathcal{E}(P) \cap \mathcal{E}(R) \subseteq \mathcal{E}(P)$, i.e., $\mathcal{E}(P) \supseteq \mathcal{E}(P \,\&\, R) = \mathcal{E}(Q)$. In brief, the assumption that $\mathcal{I}(P) \subset \mathcal{I}(Q)$ entails that $\mathcal{E}(P) \supseteq \mathcal{E}(Q)$. Since whatever entails implies, the theorem has been proved.

Remark 1 The converse of Theorem 8(i) is false, as shown by the following counterexample. Let P = Mass density, Q = Specific heat. These two are coextensive, as they apply to all bodies and only to them, but they are not cointensive. *Remark 2* Since the intension of a predicate is included in its full sense, the above theorem holds also for the latter: the richer the sense the poorer the extension. *Remark 3* In our semantics sense and reference, rather than intension and extension, occur on an equal footing: extension depends on reference and truth instead of being a basic characteristic. Moreover extension depends pragmatically, though not semantically, on sense as well. Indeed we cannot proceed to find out the extension of a predicate unless we know its meaning, i.e., its sense and reference. (Try to locate a nondescript object.) *Remark 4* To paraphrase the preceding comment: It is not that extension is a function of sense, but that a *knowledge* of sense precedes an *investigation* of extension. *Remark 5* If there were a semantic relation between intension and extension, other than Theorem 8, we would be able to determine extensions by purely conceptual means: the whole of experimental science would be unnecessary.

Theorem 8 can be given an interesting twist if rewritten in terms of the complements of the extensions, with the help of the theorem: $A \supseteq B$ iff $\bar{A} \subseteq \bar{B}$. Just as $\mathcal{E}(P)$ is the collection of objects for which P holds, so $\overline{\mathcal{E}(P)}$ is the set of objects that fail to satisfy P or, in metaphorical terms, whatever P "excludes". If P happens to have factual referents, then $\overline{\mathscr{E}(P)}$ will be the set of things or facts excluded by P. Reformulated in these terms, our last theorem becomes

COROLLARY 9.5 For any predicates P and Q of the same rank:

(i) If P and Q are cointensive then they exclude the same things;

(ii) The richer and predicate the more it excludes: If $\mathscr{I}(P) \supset \mathscr{I}(Q)$ then $\overline{\mathscr{E}(P)} \supseteq \overline{\mathscr{E}(Q)}$.

If it made sense to assign propositions (not just predicates) an extension, the previous proposition might be regarded as a formulation of Popper's idea that the more it is asserted the more is excluded. But then it would be a trivial restatement of the classical "law".

1.7. *Concluding Remarks*

We have restricted the arguments of the extension function to predicates: in the preceding theory it makes no sense to speak of the extension of a construct of a different category, such as a proposition or a theory. Save exceptions it does not even make sense to ask about the extension of the conjunction of the basic predicates of a theory, since the conjunction of predicates must be defined on a common domain (recall Ch. 1, Sec. 1.3). Thus in a group the group operation, which is binary, and the inverse operation, which is unary, cannot be conjoined to form a third predicate.

However, the restriction to predicates can be lifted in at least two ways. One of them consists in equating the extension of a proposition with that of the most complex predicate occurring in it. For example, the (strict) extension or "domain of validity" of Newton's second law of motion is the collection of bodies with size comprised between that of macromolecules and that of galaxies. Now, the law can be compressed into $\ulcorner(x)\, Nx\urcorner$, where x is the object (or referent) variable and N a complex predicate involving functions and differential operators. Hence we can set

$$\mathscr{E}((x)\, Nx) = \{x \mid Nx\} = \mathscr{E}(N)\,.$$

This extension of the theory of extensions explicates the intuitive notion of the "domain of validity" (or truth range) of a formula, familiar to scientists. The systematic development of this idea is left to the reader.

A second generalization has actually been performed by model theory but it applies only to uninterpreted formulas. The extension of an (abstract) "sentence" s is defined as the collection of models of s, i.e.

$$\mathscr{E}(s)=\{\mathscr{A} \mid \mathscr{A}\in R \,\&\, \mathscr{A}\models s\},$$

where R is the class of all relational structures of a given type. Similarly the extension of an abstract theory, such as Boolean algebra, is the set of all its models. This construal has at least two virtues. One is that it looks natural or intuitive to regard the extension of a formula, or even a whole bunch of formulas, as the totality of its realizations – provided the given formula *has* alternative realizations, i.e., it is abstract. Another is, of course, that if one does this then he gets a ready made theory of extensions, i.e. model theory. When suitably reworded this theory contains our basic Theorem 1 in Sec. 1.3 above (see, e.g., Bell and Slomson 1969 p. 159). But such a general theory of extensions is helpful only with reference to formal constructs and moreover abstract ones. As we saw in Ch. 6, Sec. 2.4, in factual science formulas are already interpreted and satisfied in some fixed mathematical structure or other, so that one is hardly interested in the totality of models of an abstract formula.

We conclude by emphasizing that our theory of extensions is not extensionalist, if only because it is based on a non-Fregean analysis of predicates. The contrast becomes most vivid in the case of a statement such as ⌜All unicorns are stars⌝. From an extensionalist point of view this is a true proposition because it is just an instance of the theorem ⌜The empty set is included in every set⌝. On the other hand in our semantics, which starts from predicates not from their extensions, "not-unicorn or star" does not refer, because its constituents "unicorn" and "star" are defined on disjoint domains. (That this is so constitutes of course a bit of empirical information.) Hence its extension is nil. Therefore the statement is false. And extensionalism does not supply an adequate account of extensions.

2. VAGUENESS

2.1. *Meaning Vagueness*

Ideally, a scientific predicate should have an exact sense, a precise reference class, and a definite extension. A predicate satisfying the first two conditions shall be called *exact*. If a predicate fails to meet either condition it will be called *inexact*. Note that exactness is not a question of extension for, in our view, truth valuations are external to constructs. Thus we may build a well organized mathematical theory concerning an unheard-of thing, a theory with an exact sense and a precise reference, but which has so far not been assigned a truth value because it has not been subjected to any tests. The peculiar predicates of this theory will then be exact even though they have been assigned no extension.

In the practice of factual science few predicates are exact. Only those belonging to a well organized theory could be exact, but sometimes they are not because of some uncertainty concerning their precise reference. A typical example of referential uncertainty is quantum mechanics, sometimes said to concern individual microsystems, at other times assemblies of such, and most often either individual systems or ensembles manipulated by observers. Under these circumstances quantum mechanical predicates are bound to be inexact in the open context of research even though they satisfy definite mathematical conditions. It is only within a precise formulation of both the mathematical formalism of the theory and its semantics, that the predicates of the theory can be exact. The applied mathematician and the mathematical physicist will not worry about interpretation problems: they will seize on the mathematical formalism shared by all the rival versions. In other words, all the quantum mechanical predicates have a hard core of meaning determined by the formalism (which includes the skeletons of the law statements) and that remains invariant under reinterpretations. This suggests introducing

DEFINITION 9.8 Let P be a predicate shared by every member T of a family $\mathcal{T}$ of theories. Then the *core meaning* of P has the following components:

(*a*) the *core sense* of P:

$$\mathcal{S}_{\text{core}}(P) = \bigcap_{T \in \mathcal{T}} \mathcal{S}_T(P);$$

(b) the *core reference class of P*:

$$\mathcal{R}_{\text{core}}(P) = \bigcap_{T \in \mathcal{T}} \mathcal{R}_T(P).$$

Our definition applies not only to the alternative interpretations of a given mathematical formalism but also to any theories that share a given predicate. For example, while the full meaning of "temperature" is determined by the totality of theories in which it occurs, their intersection determines the core meaning of that predicate. What remains outside the core is precisely the meaning vagueness of the predicate. More precisely, we adopt

DEFINITION 9.9 Let P be a theoretical predicate with given core sense $S_{\text{core}}(P)$ and given core reference class $\mathcal{R}_{\text{core}}(P)$. Then the *vagueness in the meaning* of P relative to the theory T is

$$\Delta_T\mathcal{M}(P)=\langle\Delta_T\mathcal{S}(P),\ \Delta_T\mathcal{R}(P)\rangle$$

where

$$\Delta_T\mathcal{S}(P)=\mathcal{S}_T(P)\ \Delta\mathcal{S}_{\text{core}}(P)$$
$$\Delta_T\mathcal{R}(P)=\mathcal{R}_T(P)\ \Delta\mathcal{R}_{\text{core}}(P)$$

and 'Δ' stands for symmetric (Boolean) difference.

The preceding considerations apply to theoretical predicates only. In such cases the concept of meaning vagueness turns out to be an exact concept. This is not the case with nontheoretical predicates such as "ugly". In this case one might feel tempted to try a topological approach. For example, one might wish to characterize as vague (exact) any predicate that is a limit (inner) point of a given set of predicates. But the very notion of neighborhood, necessary to define limit points and inner points, presupposes the existence of a topology of predicates. And this is forthcoming just in case the predicate family is structured, which was the case studied in Ch. 4, Sec. 2.4, but is not the case with ordinary knowledge predicates. In the latter case meaning vagueness is vague itself.

What should be done about inexact predicates? Either of two things: dressing them up in an exact garb or schooling them until they become exact. The former policy is partially implemented by allowing inexact predicates to adjust to a permissive logic of their own – e.g., some system of three valued logic (Körner 1964). We do not counsel taking this course: meaning vagueness can originate either in woolly thinking or in genuine theoretical discrepancies, and in either case it should be exhibited and worked on rather than shoved under some respectable rug. If we relax the logical standards we shall be unable to exactify our concepts within a theory as well as to discuss their differences when placed in different theories. What ought to be done is to minimize the meaning vagueness within every theory. To achieve this goal there is but one means: to improve on the logical organization and the semantics of our scientific theories. If necessary we should axiomatize them. Of course this will not guarantee the disappearance of vagueness, because alternative axiomatizations are always possible and some of them may not consist in a mere reshuffling of a fixed set of constructs. In other words, a residual meaning vagueness may be unavoidable – not as an indicator of conceptual fuzziness but rather as a healthy token of theoretical variety. While we should wish to maximize *intra*theoretical exactness we should not minimize *inter*theoretical vagueness, for this is achieved by simply outlawing all rival theories but one.

2.2. *Extensional Vagueness*

An inexact predicate is bound to be assigned an imprecise extension. For, if we are uncertain about its meaning, we shall find that there are border line cases. In this case we speak of *extensional vagueness*. A remedy for it is of course exactification. (Always go to the root of the trouble.) For example, by substituting a quantitative concept of length for the qualitative concepts "long", "medium" and "short", we get rid of meaning vagueness and at the same time decrease extensional vagueness. However, the latter may not be shrunk to nought except in very simple cases, because in general we shall have shades of truth values instead of clear cut cases of truth and falsity. Consequently extensional vagueness is sufficient but not necessary for inexactness: the former may originate in the uncertainty inherent in our truth value assignments. For this reason it is convenient to introduce a concept of extensional vagueness independent of the one of meaning vagueness elucidated in the preceding subsection.

The *strict* extension of an exact predicate P with domain D is the class of objects in D for which P is true:

$$\mathscr{E}(P)=\{x\in D \mid Px\}, \quad \text{or} \quad \mathscr{E}(P)=\{x\in D \mid \mathscr{V}(Px)=1\}.$$

The generalization to partial truth gives rise to the notion of lax extension:

DEFINITION 9.10 Let P be a predicate with domain D and $\mathscr{V}$ a truth valuation, while ε is a preassigned real number comprised between 0 and 1. Then the *extension of P to within* ε is defined as

$$\mathscr{E}_\varepsilon(P)=\{x\in D \mid 1-\varepsilon\leqslant\mathscr{V}(Px)\leqslant 1\}.$$

This concept of lax extension covers the two cases we discussed above: that of extensional vagueness due to inherent meaning vagueness, and the one due to uncertainties in truth valuation. Lax extensions include strict extensions:

$$\text{For every } 0\leqslant\varepsilon\leqslant 1,\ \mathscr{E}_\varepsilon(P)\supseteq\mathscr{E}(P).$$

The excess of the former over the latter is precisely the amount of extensional vagueness. More explicitly, we propose

DEFINITION 9.11 Let P be a predicate with domain D, lax extension $\mathscr{E}_\varepsilon(P)$, and strict extension $\mathscr{E}(P)$. Then the *extensional vagueness fringe* of P is

$$\Delta\mathscr{E}(P)=\mathscr{E}_\varepsilon(P)-\mathscr{E}(P)=\{x\in D \mid 1-\varepsilon\leqslant\mathscr{V}(Px)<1\}.$$

In words: $\Delta\mathscr{E}(P)$ includes all and only the border line cases.

The same ideas can be restated in terms of a generalized membership relation $\in_v$ defined as follows:

$$x\in_v S=_{df}\mathscr{V}(x\in S)=v, \quad \text{with} \quad 0<v\leqslant 1.$$

It is easily seen that, if $x\in_v S$, then $\neg(x\in_v S)=x\in_{1-v}S$. In terms of this generalized membership concept, the extensional vagueness fringe of a predicate P with domain D and strict extension $\mathscr{E}(P)$ becomes

$$\Delta\mathscr{E}(P)=\{x\in D \mid x\in_v\mathscr{E}(P) \quad \text{and} \quad 0<v<1\}.$$

In other words, our very sets may be allowed to be fuzzy to some extent by tying the notion of membership to that of partial truth. (For a different approach see Goguen (1969).)

Still another approach consists in focusing on statements rather than their referents. Consider an obviously vague predicate such as "healthy", or H for short, defined on the set O of organisms, i.e., $H: O \rightarrow \text{Statements}$.

Then H induces a tripartition of this set S of propositions according as the propositions Hx, with $x \in O$, are true, false, or neither. The extensional vagueness fringe of H may then be defined as the set of statements of the form Hx which prove to be neither true nor false – i.e., the statements concerning all the border line cases. If this view is adopted then the amount of extensional vagueness is given by the fraction of such alethically indeterminate statements. (Caution: do not construe such a fraction as a probability. Statements are not random variables and truth values are not allotted at random.)

We shall not pursue this matter any further but shall instead refer to the recent literature for alternative approaches (Körner, 1964; Bunge, 1967a; Gentilhomme, 1968; Goguen, 1969; Castonguay, 1972; Moisil, 1972). All of them share the conviction, laughed away by inexact philosophers, that "vagueness" can be exactified even if vagueness cannot be shrunk. There is no more paradox in this view than in the mathematical theory of approximations – nor more than in the thesis that "exact" can always be rendered more exact.

2.3. *Structural Indefiniteness*

There is a deeper rooted kind of imprecision for which no easy remedy is in sight, namely what we shall call *structural indefiniteness*. First the intuitive notion. A negative statement is less definite or less committed than an affirmative one, an existential statement less so than a universal generalization, and a possibility statement far less definite than the corresponding nonmodal proposition. Unlike the types of vagueness we have investigated before, structural indefiniteness is not due to predicate inexactness or to uncertainty in extension, but seems inherent in logical form.

A possible way of assigning indefiniteness values and thus elucidating the notion of structural indefiniteness is by adopting the following principles.

1 Atomic statements have zero structural indefiniteness.

2 The degree of structural indefiniteness of a molecular statement p equals the number of negations plus the number of disjunctions occurring in p.

3 Possibility statements, though indefinite, have no definite structural indefiniteness value.

Accordingly, for affirmative atomic statements p and q, we have

$$\begin{array}{ll} Ind(p)=0, & Ind(\neg p)=1 \\ Ind(p \,\&\, q)=0, & Ind(p \vee q)=1 \\ Ind(p \Rightarrow q)=2, & Ind(p \Leftrightarrow q)=4 \\ Ind\left(\bigwedge_{i=1}^{n} Fx_i\right)=0, & Ind\left(\bigvee_{i=1}^{n} Fx_i\right)=n-1. \end{array}$$

This concept of indefiniteness or weakness of commitment is of importance in the methodology of science (see Bunge, 1967a, I, pp. 273ff). It is often conflated with those of logical strength, content, and improbability (see Ch. 4, Sec. 3.2). Which is a reflection upon the state of semantics.

3. DEFINITE DESCRIPTION*

3.1. *The Received View: Criticism*

Definite descriptions, such as 'my mother', 'the logarithm of 1', and 'the saint next door', can pose some subtle logical and semantical problems – particularly when they refer to something that happens not to exist. It took *the* Russell to realize this and to try and analyze definite descriptions with the help of the then young mathematical logic. Russell's now classical "theory", or rather definition, boils down to this: A definite description presupposes existence and indicates uniqueness (Russell, 1905, 1919a). That is, "the so-and-so is such-and-such" must be analyzed as "There is a unique thing that is so-and-so and such-and-such". In symbols,

$$G((\imath x)\, Fx) =_{df} (\exists x)\,(Fx \,\&\, Gx \,\&\, (y)\,(Fy \Rightarrow y=x)) \tag{R}$$

Hilbert and Bernays (1968) gave a different version of the idea that "the so-and-so" is no more and no less than "there is a so-and-so and it is unique". They posited the new rule of inference

$$\frac{\begin{array}{c}(\exists x)\, Fx \\ (x)\,(y)\,(Fx \,\&\, Fy \Rightarrow x=y)\end{array}}{F((\imath x)\, Fx)} \tag{HB}$$

---

* Adapted from Bunge (1971c).

subject to certain restrictions. The differences between *HB* and *R* are these: (*a*) while *R* is a definition, *HB* is a rule of inference that, if accepted, must be added to the rules of the predicate calculus; (*b*) *R* is contextual, in the sense that the referent must be assigned a further property (i.e., *G*) to the one it exemplifies uniquely (i.e., *F*); *HB* makes no such demand but as a compensation it forces us to employ correct yet redundant expressions such as 'my mother is a mother' and 'the square of 2 is a square of 2'; (*c*) while in *R* existence and uniqueness are fused together, in *HB* they are stated separately.

Each of the above explications, *R* and *HB*, has its merits and demerits. In particular, *R* is simpler than *HB*, while *HB* involves a neat separation of existence from uniqueness. However, this detachment of the two concepts should be carried even further: one should be able to describe non-existing objects as well as objects the existence of which has not yet been established. In other words, "it is quite natural to employ descriptions *before* they have been proved to be proper" (Scott, 1967, pp. 181–182). Thus the bachelor should be allowed to say 'My wife does not exist' and the cosmologist should be able to wonder whether there is such a thing as the center of the world. Yet neither *R* nor *HB* render this possible, as both make existence a condition for definite description. For this reason both *R* and *HB* are unsuitable. The same objection holds for some other explications of the definite description (e.g., Kalish and Montague, 1957; Eberle, 1969) – certainly not for Hintikka's (1969).

In mathematics as well as in factual science questions of uniqueness are separate from questions of existence: an object satisfying a certain description may not exist or, if it does, it may not be unique. For example, the theory of differential equations contains separate existence theorems and uniqueness theorems. And in theoretical physics one can often give an unambiguous characterization of a certain object whose actual existence is far from certain: thus one may speculate on the ground state of a mesic atom that has not yet been and may never be brought into existence.

Whether in mathematics or in factual science, when attempting to validate uniqueness claims one proceeds in this way. One first assumes or exhibits existence and then one goes on to assume or investigate uniqueness under the assumption of existence: it would be wasteful to look for unique nonexistents. In other words uniqueness theorems take the form: "If there is an x with the property F then there is no y, other than x, that exemplifies F". Thus uniqueness *proofs* depend on existence assumptions or proofs but not conversely. Similarly for the empirical validation of hypotheses of factual uniqueness. But this does not mean that the *concept* of uniqueness depends upon that of existence – unless of course one is an intuitionist or an operationist. In fact the two concepts are not interdefinable. (If they were then every uniqueness theorem would be just a rewrite of some existence theorem.) And uniqueness assumptions are not deducible from existence assumptions alone.

We submit that *definite descriptions indicate only uniqueness*: that they are by themselves uncommitted as to existence – even though establishing the latter is necessary for proving the former. Otherwise the bachelor could not make jokes about his wife, the atheist could not argue about the Christian god, the physicist could not speculate about the element number 110, and the cosmologist could make no hypotheses about the center of the universe. This being so, we cannot accept any explication of definite description that involves existence. In this respect then HB is as inadequate as R. We must therefore look for a different characterization. We shall come up with two definitions, neither of which will force us to augment the set of rules of inference.

3.2. *An Elementary Analysis of Definite Descriptions*

The standard analyses of the definite description equates it with existence and uniqueness. Ours will drop existence and retain uniqueness. Now, the uniqueness condition can be expressed in a number of languages. In the present subsection we shall propose an analysis of definite description within first order predicate logic.

We are concerned with uniqueness irrespective of existence, which should be assumed or denied separately. Furthermore we wish to analyze the notion of *relative uniqueness* or uniqueness in some respect, doing it at first with the limited resources of predicate logic. There are two ways of doing this, depending on the "respect" in which an object is unique. An object may be unique because it is the sole instance of a given property, as is the case with the third power of 2. Or the object may be unique because of a relation it bears to some other object, as is the case with my mother. We shall elucidate these two notions by means of the following conventions.

DEFINITION 9.12 The object a is unique in the respect denoted by the predicate $F =_{df}$ a exemplifies F and there are no further individuals, other than a, that exemplify F:

$$a \text{ is } F\text{-unique} =_{df} Fa \;\&\; \neg(\exists x)(x \neq a \;\&\; Fx)$$

or, equivalently,

$$a \text{ is } F\text{-unique} =_{df} Fa \;\&(x)\,(Fx \Rightarrow x = a).$$

Next we stipulate that "a is F-unique" amounts to "the x such that x is an F". More explicitly, we lay down

DEFINITION 9.13 The object a is (equal to) the x such that x is an F, just in case a is unique:

$$(a = (\lrcorner\, x)\, Fx) =_{df} a \text{ is } F\text{-unique}.$$

We have chosen '$\lrcorner$' to designate the definite descriptor both for typographic convenience and to forestall confusion with Russell's symbol, which designates a different concept.

The above definitions entail our *elementary explication of the definite description*:

$$(a = (\lrcorner\, x)\, Fx) =_{df} Fa \;\&(x)\,(Fx \Rightarrow x = a) \qquad (N)$$

For example, $\ulcorner 2$ is the smallest prime$\urcorner$, i.e., $\ulcorner 2 = (\lrcorner\, x)\, SPx \urcorner$, is now analyzed as: $\ulcorner SP2 \;\&(x)\,(SPx \Rightarrow x = 2)\urcorner$, where in turn SPx is defined as $(y)\,(Px \;\&\; Py \;\&\; x \neq y \Rightarrow x < y)$.

For N to hold Fa need not be true. And if Fa is not asserted (separately) then $(\exists x)\, Fx$ does not follow. Consequently N does not involve (entail) existence. For example, the equality

Zeus = The boss of the Greek Olympus.

which is a substitution instance of the left hand side of N, does not commit us to paganism. It is just a designation convention. If challenged, then the right hand side will be challenged as well but the definitory equality will still hold. Likewise if N is construed as an equivalence, since for $\ulcorner A \Leftrightarrow B \urcorner$ to hold both sides must have the same truth value, e.g., falsity. If we now assert (separately) that Zeus is a boss of the Greek Olympus then we conclude that Zeus exists; and if we deny the same statement, i.e., negate *Fa*, then we commit ourselves to the statement that Zeus does not exist.

So much for definite descriptions in terms of unary predicates. We now generalize our analysis to predicates of any rank. But to keep the exposition readable we limit our definitions to binary relations.

DEFINITION 9.14 The object a is unique in its relation R to $x =_{df}$ a bears the relation R to x and there are no other individuals y, except a itself, that are R-related to x:

$$a \text{ is } R\text{-unique in its relation to } x =_{df} Rax \,\&(y)\,(Ryx \Rightarrow y = a).$$

DEFINITION 9.15 The object a is (equal to) the x such that x bears the relation R to $b =_{df}$ a is unique in its R relation to b:

$$\begin{aligned}(a = (\lrcorner x)\, Rxb) &=_{df} a \text{ is } R\text{-unique in its relation to } b \\ &=_{df} Rab \,\&(x)\,(Rxb \Rightarrow x = a).\end{aligned}$$

To sum up, we have identified definite description with uniqueness. Unlike the accepted view ours does not involve the assumption that the described individual exists in some context: it only asserts the *non*-existence of *other* individuals satisfying the same conditions. This is as it should be, for existence is not a matter of definition, let alone of designation: existence is a matter of either assumption or validation. In other words an existence claim is a hypothesis that must be substantiated not a convention that can be freely introduced. For example, whether or not a function with given properties exists, is a matter for an existence theorem to decide. And whether or not a concrete object with assumed properties exists, is a matter that cannot be decided without performing empirical tests. These methodological requirements are violated by the construals *R* of Russell and *HB* of Hilbert-Bernays (see Sec. 3.1) but not by ours (*N*). Indeed, according to our construal the existence of the described individual may, but need not, be asserted separately, namely thus: $(\exists x)$ (x is F-unique), or thus: $(\exists x)$ (x is unique in its R relation to b). Consequently there will be no *logical* differences between a proper (or full) description and an improper (or empty) description: the differences are solely semantic. The semantic peculiarities of the definite description will be taken up in Secs. 3.5 and 3.6 in the light of a deeper yet simpler analysis of our subject, to which we turn presently.

3.3. *A Mathematical Analysis of Definite Descriptions*

We shall now avail ourselves of the general concept of a function, which goes beyond the predicate calculus. Consider the formula ⌜The cost of x is (equals) y⌝ or, more briefly, ⌜x costs y⌝. For every substitution instance of x there is exactly one value of y such that y equals the cost of x. Hence cost is a function – call it C – so that we can write: ⌜$C(x)=y$⌝. Likewise ⌜The father of x is y⌝, or shorter ⌜y fathered x⌝, can be symbolized as: ⌜$F(x)=y$⌝, where 'F' stands for the fatherhood function. These symbols convey the idea that cost and fatherhood are properties of something and moreover that these properties are adequately represented by functions in the mathematical not in the logical sense, for their values are not statements but further individuals. These examples may also be construed as relational statements with the qualification that the relations C and F be many-one. But this construal, suitable for ordinary knowledge cases such as ⌜Scott wrote *Waverley*⌝, is inadequate for most scientific purposes. In science one prefers functional statements of the form "$F(x)=y$".

Let us now *truncate* the functional formula ⌜The F of x equals y⌝ by leaving out the value of the function. We thus get "The F of x", or "$F(x)$", which may be called a *functional semistatement*. This expression indicates the function of interest and an arbitrary value of its argument but not the corresponding value of the function. If the function does have a value at x, i.e., if F is "defined" at x, then that value is unique by definition of the concept of a function. And this is all a definite description indicates, namely a unique object. We compress the preceding into

DEFINITION 9.16 Let F be a function from a set A into a set B. Then the expression '$F(a)$' is called a *proper definite description* $=_{df}$ F is defined at $a \in A$.

If $F(a)$ is a proper definite description then it names a unique individual, say b, in the codomain B of F. That is, the relation between a proper name b and a definite description $F(a)$ is now the full fledged statement "$b=F(a)$", read 'b is the F of a'. In this case $F(a)$ is a name and it poses no further problems. Otherwise we have an improper or empty definite description. More explicitly we lay down

DEFINITION 9.17 The functional semistatement $F(a)$ is an *improper definite description* $=_{df}$ F is not defined at $a\in A$.

Examples: "The weight of my thoughts", "The father of the universe". Suggestion: a number of metaphors are just improper descriptions. This remark might be useful in analyzing the structure of some metaphors.

An alternative but essentially equivalent explication is obtained with the help of the concept of a partial function, or a correspondence between a *sub*set of A and a set B. Thus "king" and "president" are partial functions on the set of countries: "king" is a total function on the set of monarchies and "president" a total function on the set of republics. In general we have

DEFINITION 9.18 Let F be a partial function with domain A. Then $F(a)$ is a *proper (improper) definite description* $=_{df}$ a belongs (does not belong) to A.

On either of the two functional construals, the description can be stated and analyzed without even having to introduce a special symbol. The link between the preceding elementary elucidation (Sec. 2) and the present one is supplied by

DEFINITION 9.19 Let P be a unary predicate with domain B and let F be a function from a set A into B. Then for all $b\in B$

$$b=(\text{℩}x)\,Px=_{df}F(a)=b.$$

Given a unary predicate P it is always possible to find the corresponding function F that will satisfy the preceding convention and conversely. For instance "golden mountain" may be construed not only as a (molecular) predicate but also as a function on the set of places. Consequently assuming that the name of the golden mountain is 'Shiny', Definition 19 gives

> Shiny = The golden mountain $=_{df}$ The golden mountain at place a = Shiny.

There is more to the relation between predicates and functions: let us look into it.

3.4. *Continuation of the Analysis*

Take “talented” as an example of a unary predicate P and “author of” as an instance of a function F. (We brush aside the case of joint authorship.) Both P and F are mappings: P maps writers into statements while F maps books into writers. That is,

$$\begin{array}{ll} P & \text{Talented: Writers} \rightarrow \text{Statements} \\ F & \text{Author of: Books} \rightarrow \text{Writers} \end{array}$$

In the first case we have, e.g., ⌜Walter Scott is talented⌝ while in the second we may have ⌜The author of *Waverley* is (equal to) Walter Scott⌝. (The *is* in the first statement is predicative while the *is* in the second statement is that of equality.) This allows us to write

$$A(w) = s,$$

where '$A(w)$' stands for “the author of *Waverley*” and 's' for Walter Scott. Consequently ⌜Walter Scott is talented⌝ can be transformed into ⌜The author of *Waverley* is talented⌝, i.e., briefly $T(A(w))$. Now, this is just an instance of the composition of the functions A and T, i.e.,

$$\text{Books} \xrightarrow{\text{Author of}} \text{Writers} \xrightarrow{\text{Talented}} \text{Statements}.$$

This composition may be represented as a commutative diagram.

$$\begin{tikzcd}
\text{Books} \arrow[r, "\text{Author of}"] \arrow[dr, "ToA"'] & \text{Writers} \arrow[d, "\text{Talented}"] \\
& \text{Statements}
\end{tikzcd}$$

Obviously this works only when the functions concerned are defined at all the points of their domains and when the range of the first equals the domain of the second. This remark will help us to solve a problem concerning the relation between definite descriptions and names.

As Russell pointed out, definite descriptions are not names. They may have the same referent but they do not mean the same. Thus 'The mother of God' "says" more than 'Mary' in the context of Catholic theology. Whatever significance a name may be assigned (and this is a controversial point), the associated definite description, if any, must have a sense containing the previous one. Nevertheless for the purposes of deduction proper definite descriptions can be treated just like names. For example, 'the sum of 2 and 3' can be replaced by the numeral '5'. In other words, although a definite description *is* not a name it behaves in inference *as if* it were a name – provided the necessary precaution is taken. The condition for this functional or behavioral identity of names and definite descriptions is, of course, that the function F in question be defined at the point of interest. For only in this case will $F(a)$ constitute a proper or nonempty definite description according to Definition 16. In other words, a definite description can be treated as a name provided there *exists* an individual complying with the description. Example:

$$\begin{array}{rl} Ts & 1 \\ A(w)=s & 2 \\ \therefore T(A(w)) & \text{1, 2, Principle of identity.} \end{array}$$

This inference is valid because, on the interpretation given above, the function A is defined at w: indeed, at w, A takes on the value s. On the other hand if now 'w' stands for the world, then in the context of a naturalistic world view the above inference is invalid no matter how 'T' be reinterpreted, for A is no longer defined at w. In short, with this new interpretation of 'w' $A(w)$ becomes an improper description and the substitution of $A(w)$ for s is invalid because one of the conditions for the composition of functions, i.e., for the existence of the composite $T \circ A$, is no longer met. (We may, if we wish, say with Frege that '$A(w)$' now designates the null individual, but this won't save the inference.) In this regard, then, our treatment does not differ from either Russell's or Hilbert-Bernays'. The difference consists, of course, in that the existence condition is now stated separately instead of being fused with the definite description.

3.5. *Meaning Questions*

We know from Ch. 7 that meaning is contextual. Thus "The creator of the universe" is meaningful in some theodicies but not in physics, where the concept of creator does not occur. Hence we had better start by recalling the notion of a context (Ch. 2, Sec. 3.4, Definition 10). We shall reword it in terms of functions rather than predicates. *Voilà*: The ordered triple $\mathbb{C}=\langle S, \mathbb{F}, U\rangle$ is called a *context* iff S is a set of statements in which only the function constants in the function family $\mathbb{F}$ occur, and the reference class of every F in $\mathbb{F}$ is included in the universe U.

We now postulate the conditions under which a definite description, construed as a functional semistatement (Sec. 3.3), makes sense in a given context and has a referent in it.

AXIOM 9.1 Let $F(x)$ be a definite description and let $\mathbb{C}=\langle S, \mathbb{F}, U\rangle$ be a context. Then $F(x)$ *makes sense* in $\mathbb{C} =_{df} F$ is in $\mathbb{F}$.

AXIOM 9.2 Let $F(x)$ be a definite description and let $\mathbb{C}=\langle S, \mathbb{F}, U\rangle$ be a context. Then $F(x)$ *has a referent* in $\mathbb{C} =_{df} F(x)$ makes sense in $\mathbb{C}$ and F is defined at $x \in U$.

Finally we identify the referent in an unambiguous way:

DEFINITION 9.20 Let $F(x)$ be a definite description that makes sense and has some referent in a context $\mathbb{C}=\langle S, \mathbb{F}, U\rangle$. Then the value $y=F(x)$ is called the *referent* of $F(x)$.

Notice that the above axioms are postulates not definitions: indeed they do not define sense and reference but just stipulate the conditions for a definite description to have them. Now, in our semantics, for an expression to be significant it must have a (vicarious) sense. (Recall Ch. 7.) It follows that "The present king of France" and other referentially vacuous definite descriptions may be meaningful in certain contexts. In other words, an empty description may stand for a concept, i.e., something endowed with a sense, even though it fails to denote, i.e., to have a real counterpart. Consequently ⌜The present king of France is bald⌝ is meaningful as well: it has a sense – and even a referent. Whether this referent exists, i.e., whether the statement is true, is another matter – one to be considered shortly.

Our theory is at variance with the view that empty descriptions and sentences containing them "suffer from infelicity" because devoid of real reference (Austin, 1962). This opinion would be justified if meaning were identified with extension. But, as we have seen in Sec. 1 and before, such a referential view of meaning would be crippling to science, where statements about "theoretical" (i.e., hypothetical) entities have to be made, discussed and put to the test long before it becomes possible to ascertain whether they have any real referents.

Another advantage of our theory is that it dissolves a well known puzzle about "intensional expressions", i.e., non truth functional formulas. Consider the descriptions

The square root of 4. (*A*)
The atomic number of Helium. (*B*)

Since *A* and *B* have the same designatum, namely 2, they are equivalent. And, being equivalent, they should be exchangeable to a nominalist (e.g., Ajdukiewicz, 1967a). But of course the two descriptions have different referents: while *A* concerns 4, *B* is about Helium. Therefore *A* and *B* are not synonymous. Consequently, on our theory they are not everywhere exchangeable, whether in "extensional" (i.e., truth functional) or "intensional" (i.e., non truth functional) contexts. Hence the statement

Archimedes knew that the square root of 4 equals 2. (*C*)

is not the same as

Archimedes knew that the atomic number of Helium is 2. (*D*)

3.6. *Truth Questions*

Consider again the much discussed statement ⌜The present king of France is bald⌝. Can we say that it is false or that it is true? Russell held it to be false and most philosophers seem to share his view for the simple (ergo suspect) reason that it is a proposition and thus (allegedly) either true or false and, not being clearly true, it must be false. However, some philosophers have felt dissatisfied with this view. Thus Frege, and at one time Strawson (1950), held that statements containing empty descriptions are neither true nor false. More recently Strawson (1964) has come to the conclusion that his former stand is not compelling: that each side has its merits and that it is unimportant which is taken.

In our view truth and falsity do not inhere in propositions but are (sometimes) attributed to them (Ch. 8). Now, in order for a statement containing a definite description to be assigned a truth value in a given context, it must point to a definite referent in some context. For, if it fails to have a referent, then the statement cannot be "faced" to it in order to be assigned a truth value. Therefore we stipulate

AXIOM 9.3 A statement containing a definite description $F(x)$ may be assigned a truth value in the context $\mathbb{C} = \langle S, \mathbb{F}, U \rangle$ iff $F(x)$ has a referent in $\mathbb{C}$ and this referent exists.

For example, "the luminiferous ether" is a definite description whose sense may be taken to be determined by some ether theory (recall Ch. 5). In all mechanical ether theories, such as Cauchy's, the statement

$$e = \ulcorner\text{The luminiferous ether is elastic}\urcorner.$$

is not only meaningful but also true. However, since the predicate "ether" is absent from modern optics, by our Axiom 1 (Sec. 3.5) it makes no sense in it. Consequently the definite description "the luminiferous ether" makes no sense in modern optics. Ergo no statement containing this definite description makes sense in modern optics: in particular, e above is not meaningful in modern optics, or *MO* for short. And, not being meaningful in *MO*, e cannot be assigned a truth value in this context. In other words, the truth valuation function $\mathscr{V}$, that may be construed to carry the couples statement-context into truth values, is not defined for the pair $\langle e, MO \rangle$. In short, $\mathscr{V}$ has no value at $\langle e, MO \rangle$ even though $\mathscr{V}(e, E) = 1$, where 'E' is short for "ether theory". (Nor does e take the value "indeterminate", as some interpretations of many valued logic would have it. A function not determined, or not defined, for a certain value of its argument has no value at it. Depending on the theory of truth $\mathscr{V}$ may take two or more values but it cannot take the value "indeterminate".) What holds for the ether holds for contemporary French kings: in the context of contemporary history ⌜The present king of France is bald⌝ is neither true nor false. (It only presupposes a false proposition.) So let us stop arguing about it.

Finally note that the Liar paradox does not occur in our semantics, because 'What I am now saying' is a definite description not a proposition, hence it cannot even be false.

3.7. *The Real Size of the Theory of Descriptions*

The analysis of definite descriptions has been overrated to the point of having been regarded as Russell's chief contribution to philosophy – which is a way of belittling *PM*. On the other hand the theory of descriptions has been underrated and even misplaced by many ordinary language philosophers. Some took it to be a matter of literary criticism, others took it for a grammatical analysis of the definite article – as if 'my wife', 'Plato's teacher', and 'that weird chap in the corner' did not qualify as definite descriptions. As has been pointed out, all advanced languages teem with definite descriptions even if, like Latin, they lack definite articles. Also mathematics, the language of science, is shot through with definite descriptions – just recall 'the sine of 10°', 'the composition of f and g', and 'the indefinite integral of f'. So is contemporary science, which has more use for definite descriptions – e.g., in the form of spatio-temporal coordinates – than for proper names. This renders Quine's proposal of assimilating all names to definite descriptions attractive. However, (*a*) even if in practice we often do proceed this way, it is convenient to have the complex notion of description analyzed into simpler concepts, and (*b*) in metaphysics we need the concept of a non-descript or bare individual that may function as one of the building blocks of the notion of a fully qualified thing. (See Ch. 1, Vol. 3 of this *Treatise*.)

Since definite descriptions are rampant, it behooves the philosopher to analyze them. But the analysis need not introduce any new technical concept: we have seen that descriptors are reducible to standard components of elementary logic and mathematics. Our evaluation of description theory lies thus midway between the two views that are currently dominant: instead of either renouncing definite descriptions or inflating them, we hold them to be normal constituents of any language with a reasonable expressive power. Moreover in our view the syntax of definite descriptions is trivial: only their semantics is somewhat complex, in the sense that it involves the notions of sense, reference, and truth. But then this is what semantics is all about: sense, reference, and truth.

This is as far as we go in applying our basic doctrines to issues in pure semantics. The next chapter, which is also the last, explores some of the relations among semantics and its neighbors.

CHAPTER 10

NEIGHBORS

In this, the final chapter, we shall peek at some of the adjacent fields of inquiry in order to better locate ours. We shall try and see what they look like in the light of our semantic candle. In each case we shall have to limit ourselves to examining a few typical problems. Moreover our discussion will be rather quick, as our aim is to explore the nature of the bonds between the semantics of science and its nearest neighbors rather than to examine the latter in detail. We shall first glance at mathematics, or rather its philosophy, then at three traditional branches of philosophy: logic, epistemology, and metaphysics.

1. MATHEMATICS

1.1. *The Relevance of Semantics to Mathematics*

That mathematics is relevant to exact semantics, is analytically true, for exact semantics is nothing but semantics built *more geometrico.* The question is whether basic semantics is relevant to mathematics – whether the former can add anything to the semantics of mathematics, or model theory. And this is not obvious.

Consider the notions of designation, reference, sense, and truth – all of which would seem to be of interest to mathematics. (The concept of model, in the model theoretic sense, is the exclusive property of the semantics of mathematics and it does not seem portable to other fields: recall Ch. 6, Sec. 2.4.) The concept of designation is almost trivial and so is that of reference in the case of mathematics – though not so in relation to science. Indeed, to all but the literalists it is clear that mathematical symbols designate mathematical constructs. It is equally plain that mathematical concepts and statements either are or concern mathematical objects, which are so many constructs. Thus while "two" does not refer to anything, "even" refers to integers and ⌜Two is an even number⌝ refers to two, which can in turn be designated by the numeral '2'. So far nothing shattering.

As to the third generic semantic concept, namely that of sense, we do have something to say about it, particularly since model theory deals only in extensions. (Recall Ch. 6, Sec. 2.3.) But in our view the full sense of a theory, whether mathematical or factual, is ultimately determined by the postulates of the theory. (See Ch. 5, Sec. 5.) Hence we have nothing to add to what was said in Chapters 4 to 7. (For specifics concerning meaning in mathematics see Castonguay (1972).) However, we may caution against the persuasive or ideological uses of 'sense' and 'meaning' as exemplified by the war-cry that whatever fails to be constructive, such as the axiom of choice, is meaningless (cf. Lorenzen, 1967). This Principle of Intolerance is meaningless unless backed by a definite theory of sense (or of meaning) in the spirit of constructivism – which, alas, is not available. So much for sense in mathematics.

On the other hand we have little to say about the fourth capital term of semantics, namely 'truth', with reference to mathematics. The reason is that, although 'truth' parades as a general term, it is not: 'truth' is an ambiguous term designating two radically distinct concepts, namely those of formal truth and factual truth. Whereas factual truth is elucidated in terms of external reference and empirical evidence, formal truth is elucidated in terms of satisfiability and proof. (From a pragmatic point of view proof has the upper hand, since showing that a formula is true in a certain model, or under a certain interpretation, boils down to proving the formula in the theory of the model.) Whereas factual truth is the object of a special theory, formal truth is elucidated within model theory, which is a multiple purpose theory. (Also the notion of potential formal truth is definable in model theoretic terms: see Robinson, 1965. And even the notion of partial formal truth can be handled by model theory: see Chang and Keisler (1966).) Because our system of semantics is geared to factual science, it is as irrelevant to mathematical truth as the model theoretic concept of truth is alien to factual truth.

In sum, only our theory of sense (Chapters 4 and 5) makes sense with reference to mathematics. Having nothing further to offer we shall close with a couple of critical remarks.

1.2. *On Extensionalism*

'Extensionalism' is sometimes taken to designate the thesis that ordinary logic is the only logic one needs – that, in particular, modal logics are dispensable. We accept this thesis but reject that designation as mistaken and misleading. (See Ch. 4, Sec. 1.3.) The authentic extensionalist thesis is, in a nutshell, that *every concept worth its name is a set.* When adopted in semantics, extensionalism effects a hecatomb: it kills sense (in every sense) and it is apt to conflate reference with extension.

The prestige of the extensionalist thesis stems from the belief that it has subdued mathematics. In fact it is widely held that (*a*) set theory is fully extensional and (*b*) the whole of mathematics is reducible to set theory. However, these two tenets are, to say the least, controversial. First, set theory contains a basic concept, that of membership, which is not construed as a set but as a relation between something (whether individual or not) and a set. Not only is the membership relation not defined as a set (of ordered pairs), but the latter is partially defined in terms of the former, namely by saying that, if $x \in y$, then y is a set. (But of course only the totality of postulates of some set theory does the job of determining the full sense of $\in$.)

Second, set theory contains a postulate, the principle of separation (or its forerunner the principle of abstraction), that relates predicates to sets – the extensions of the former – without defining the former in terms of the latter. (Recall Ch. 4, Sec. 1.2 and Ch. 9, Sec. 1.1.) If this were the sole assumption of every set theory then it would support Russell's view that reasoning on properties is primary – and by the same token it would disqualify Bourbaki's Olympian contempt for the allegedly anachronistic *raisonnement en compréhension.* However, a balance between the two extremes looks more realistic than either.

Third, although nowadays nearly every mathematical theory – much to Wittgenstein's chagrin – uses set theoretic concepts and even some formulas of set theory, it must add something of its own if it is to get off the ground. Without such specific concepts and assumptions, that are not reducible to (definable in or deducible from) set theory, there would be no mathematical theories other than set theory. (The fact that most new mathematical concepts can be characterized with the *help* of set theoretic concepts does not entail that they were already "contained" in set theory. Likewise an organism is not prefigured in its physical components.) Thus the definition of one of the simplest mathematical structures, namely a semigroup, requires the notion of associativity, which set theory does not define.

The upshot of the preceding discussion, conjoined with that in Ch. 4, Sec. 1.2, is that the extensionalist thesis is false in mathematics. On the other hand the program of using set theoretic concepts in the whole of mathematics and its applications has been tremendously fruitful although it may not be the last word. In any case "set theorification" should be kept distinct from extensionalization.

1.3. *On Objectivity*

Our second and last remark will concern objectivity. It is almost universally agreed that mathematics is objective – but this because everyone seems to have his own concept of objectivity. In any event this much seems to be true: Although mathematics is a creature of the brain, it is not in the same category with (wild) dreams and (silly) fairy tales. It does not hold by arbitrary decree or because I want to or because we happen to believe in it. Once born, a piece of mathematics ceases to be subjective and acquires a certain objectivity. In fact it is as objective as factual science – only in a different sense. Let us compare these two modes of objectivity.

The objectivity of factual science consists in its reference to external objects alone (semantic objectivity) and in its impersonal or publicly scrutable checking procedures (methodological objectivity). Even mental processes, when studied by scientific psychology, are handled as external objects and in a publicly criticizable way. Mathematics, on the other hand, fails to refer to external objects, whether ideal or material, hence it is not semantically objective. Nor is it semantically subjective: it is not concerned with our private mental states. The objective/subjective dichotomy, in the semantic sense, applies to mathematics no more than the hot/cold dichotomy does.

But mathematics is *methodologically objective*, though not in the sense that it employs empirical testing procedures (observation, measurement, experiment). The methodological objectivity of mathematics consists in (*a*) impersonality, (*b*) abidance by the assumptions and rules agreed on beforehand – including the general principles of rational argument, and (*c*) the justification of both assumptions and rules in terms of impersonal values such as coverage, systemicity, validity, and clarity. All three traits are shared with factual science, which adds empirical testing to them. Hence mathematical objectivity is but a special case of scientific objectivity: objectivity without objects other than mathematical constructs.

In sum, mathematics does not possess the semantic objectivity that both Platonists and vulgar materialists assign it: mathematics is neither about self existent ideas hovering over the world nor about the latter. Mathematics is methodologically objective in the sense that its procedures are exoteric. But mathematics is not semantically objective. To claim that it is (as Popper (1972), (1974)) does, i.e. to hold that mathematics is just as objective as physics, is to conflate epistemological realism with objective idealism.

2. LOGIC

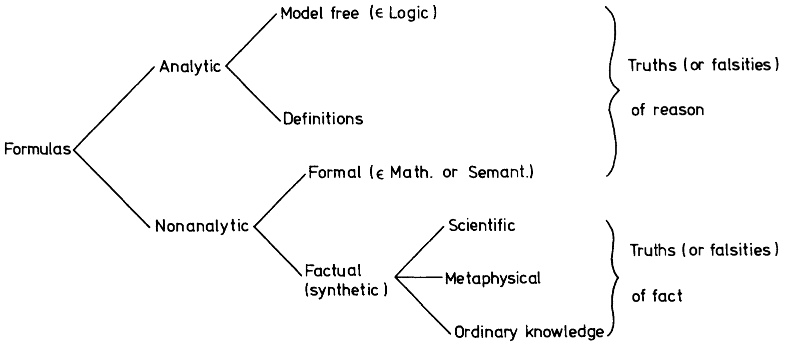
2.1. *Analyticity*

The problem of analyticity has held the attention of philosophical semanticists to the detriment of all other semantical problems, thus creating a serious imbalance and causing boredom. So far in this work we have made a modest use of the concept of analyticity, or rather one of them, without elucidating it. The time is ripe for such an elucidation. As usual we must choose the target before shooting: we must decide whether we want to have a narrow concept or a wide one. Having tried a concept of analyticity of maximal extension in the past (Bunge, 1961b) we shall now choose the most restricted one. The strict (and relative) acceptation of 'analytic' is this: An analytic formula is one that either holds under all interpretations (in all models) or consists in a definition. More explicitly, we adopt the following

DEFINITION 10.1 A formula ϕ in a theory T is *analytic in* $T =_{df}$ ϕ is either a definition in T or model free.

This convention does not employ the notions of logical form and meaning but it fits the statements that are (formally) true "by virtue of their form", i.e. the tautologies, as well as those that hold "by virtue of the meanings of their parts" – e.g. the dictionary definitions (which Carnap called 'semantic postulates').

To call 'synthetic' all those formulas that are not analytic would be misleading in view that 'synthetic' has so often been equated with 'having a factual content' or with 'empirically testable' or with 'informative'. The extralogical formulas of a mathematical theory are not analytic in the sense of Definition 1 but it would be queer to call them 'synthetic': they are *neither* analytic nor synthetic (in the sense of empirical or factual). In other words, *the analytic/synthetic distinction is not a dichotomy.* We must distinguish more than two species of statement, at least the following ones:


Formulas
Analytic
Model free (∈ Logic)
Definitions
Nonanalytic
Formal (∈ Math. or Semant.)
Factual (synthetic)
Scientific
Metaphysical
Ordinary knowledge
Truths (or falsities) of reason
Truths (or falsities) of fact

The preceding definition of "analyticity" seems to be precise and clear but it does not solve the practical problem of spotting analyticity in every single case. Such a problem will come up in ill organized conceptual bodies such as those of ordinary knowledge and of intuitively formulated scientific theories. For example, the statement ⌜Seeds germinate when they fall on fertile soil⌝ may be taken either as a lawlike statement or as a covert (and partial) definition of soil fertility. If ambiguities like this arise in the flabby context of ordinary knowledge, so much the worse for the latter – or for the attempt to draw technical distinctions in nontechnical contexts. No such status ambiguities occur in a reasonably well organized conceptual body. What does happen in the latter case is that one and the same formula may be analytic in one systematization and nonanalytic in the next – but we made room for such switches by relativizing the concept of analyticity to that of theory in Definition 1.

In conclusion, our definition of analyticity solves the theoretical problem of elucidating this notion but is not a fool proof *criterion* for telling whether a given formula is analytic – much less for identifying analytic components in open contexts. (Likewise a definition of consistency is not enough to prove the consistency of any particular theory.) Nevertheless a difficulty in drawing a distinction in particular cases does not prove that there *is* no difference. Nor does it prove that it is impossible to reorganize the body in question so as to bring out the difference, not just between analytic and nonanalytic statements, but also between hypotheses and conclusions, and so on.

If the preceding considerations are substantially correct then the concept of analyticity is far less important to philosophy than that of truth of reason or necessary truth (Leibniz, 1714). While this wider concept serves to distinguish formal science from factual science, the narrower one of analyticity serves to characterize logic *vis-à-vis* the rest and, in particular, mathematics. Indeed, while the two formal disciplines contain only necessary truths, logic has the monopoly of analytic formulas other than (extralogical) definitions. Mathematics does not get started unless some nonanalytic (and nonfactual) formulas, containing extralogical concepts such as "$\leqslant$" and "$+$", are added to a basic logical broth. For example, the theory of groupoids adds the following ingredients to the logical predicates: an abstract set and a binary operation on it.

In sum, given a set R of rules of inference, logic is self-generating, i.e. the totality of analytic formulas follows from no assumption whatever. (The catch is in R.) On the other hand a piece of mathematics requires, in addition, a nonempty set of extralogical (but also nonfactual) assumptions. In symbols: while $\emptyset \vdash_R Logic$, $A \vdash_R Mathematics$, where A is the set of mathematical assumptions. (A change in R may shift the border line without obliterating it.) An equivalent characterization of the difference is this: While the truths of logic are satisfied in all models, those of mathematics are satisfiable only in some models – sometimes just one, at other times infinitely many, but never all.

We uphold then the analytic/synthetic distinction, indicted in recent times (Quine, 1952). However, we do not define analyticity in terms of the information needed for understanding a sentence, this being a pragmatic not a semantic concept of analyticity; hence we are not bothered by Quine's examples. Furthermore we do not regard the distinction as a dichotomy or as being central to the whole of semantics and the philosophy of formal science. The essential distinction, as far as epistemology and the philosophy of science are concerned, is that between truths (or falsities) of reason and truths (or falsities) of fact. Were we to give it up we might be tempted to indulge in either of the following oddities: empiricism with regard to formal science, and apriorism with reference to factual knowledge. To show that these are not imaginary perils we quote an instance of each. The empiricist philosophy of mathematics is defended by no less than Kalmár (1967 criticized by Goodstein 1969). And the converse move, of eliminating extralogical postulates in favor of definitions, is not only a favorite ploy with conventionalists (criticized by Enriques, 1943, pp. 250–251) but has also been tried by Quine and Goodman (1940, Quine, 1964, criticized by Bunge, 1967a, I, pp. 132–133). Finally, many mechanics textbooks contain traces of Mach's attempt to blend the two moves – taking law statements for definitions and conversely (see Bunge, 1966). If only to avoid such errors it is imperative to hold the fort – not the analytic/synthetic nondichotomy but the rational/factual dichotomy. Still, it is equally imperative (*a*) not to insist on drawing such metatheoretical distinctions with regard to conceptual bodies that have not been given a deductive structure, and (*b*) not to forget that pure reason is an invention of certain organisms.

We conclude by sketching an enlarged notion of semantic analyticity, or analyticity by virtue of meaning. This notion is suggested by our incursion into the topology of the intension space and, in particular, by our elucidation of Wittgenstein's coarse notion of family resemblance (Ch. 4, Sec. 2.5). The idea is that an exact tautology can be approximated, arbitrarily closely, by a *quasitautology* or statement that is *almost* completely true (formally). Consider ⌜All A's are A's⌝ and replace one of the occurrences of A by B, where B is a close relative of A in the sense that B belongs to a small neighborhood of A in accordance with Definition 9 in Ch. 4, Sec. 2.5. Then both ⌜All A's are B's⌝ and ⌜All B's are A's⌝ will be *quasitautologies*. The process is reversible: given a nontautologous statement, Theorem 4.11 in Ch. 4 allows us to construct a whole sequence of propositions the limit of which will be a tautology. Take for instance the much debated statement ⌜The fittest organisms survive best⌝. Replace "fittest" by "most fertile" (in accordance with the Pickwickian acceptation adopted by population genetics) and "survive" by "propagate" (or "have the best reproductive survival") and you get an exact tautology. In sum, if meaning is a matter of degree so is analyticity. And, just as in the case of the multiplicity of truth values, the continuum of degrees of analyticity is consistent with rigid two valued logic.

2.2. *Definition*

The concept of a definition has proved more problematic than it deserves (see e.g. Bunge, 1967a, I, pp. 117–139). This is partly due to the adoption of too wide a construal of 'definition' – one that makes room for practically every determination of one group of constructs by another. A number of problems drop out if we adopt a narrower concept – just as with the case of analyticity. The narrowest of all, and that which we favor, is Peano's construal of every definition as an equality of the form: "*the defined = the object defining*", where the object in question is either a sign or a construct (Peano, 1921).

Example 1 The definition of set equality: ⌜If A and B are sets, then $A=B=_{df}$ The membership of $A=$The membership of B⌝. *Example 2* The definition of "$\leqslant$" in the relational structure $\mathcal{N}=\langle N, +\rangle$, where N is the set of natural numbers and $+$ the addition operation: ⌜If x, y and z are natural numbers, then: $x\leqslant y=_{df}(\exists z)(x+z=y)$⌝. *Example 3* The axiomatic definition of a groupoid: ⌜The structure $\mathcal{G}=\langle G, \cdot\rangle$, where G is a nonempty set and $\cdot$ a binary operation on G, is a *groupoid* $=_{df}$ For all x and y in G, $x\cdot y$ is in G⌝. *Example 4* The definition of the energy density of an electric field: ⌜If E represents the intensity of the electric electric field f, then $U(f)=_{df}(1/8\pi)\,E^2$⌝.

Note the following characteristics of a definition, whether explicit or implicit, conditional or unconditional, by abstraction (as Example 1) or axiomatic (as Example 3) or of any other type provided it is centered on the equality concept. Firstly, it may be taken to define a *construct or the symbol* chosen to designate it. One does not define things other than constructs or their symbols. For example one does not define light (but describes it); rather, one defines some concept of light (or the terms 'light', 'lux', 'lumière', etc.). More on this toward the end of this section. Secondly, every definition is *relative* to some context, whether a relational structure (as in Example 2) or a full theory (see Padoa, 1901 and Tarski, 1934). The contextual or relative character of definitions should be kept in mind in order to avoid some of the mistakes mentioned at the end of the preceding subsection and if one wishes to understand why one and the same word, when defined in different contexts, may acquire different significations – i.e. may designate different constructs. Thirdly, the concept "$=_{df}$" of equality by definition should be assigned all the formal properties of *equality* (or congruence), in particular the symmetry property, in order to ensure the exchangeability of definiens and definiendum. This requirement is less trivial than it looks for, from a methodological point of view, there is no such symmetry: the definiendum and the definiens cannot exchange places. In other words, from a pragmatic point of view '$=_{df}$' is a metalinguistic symbol that intends to convey the idea that the left hand side is determined by (is a function of) the right hand side not conversely. (This is how definitions were treated in *Principia Mathematica.*) However, this difference between definiens and definiendum is metatheoretical and methodological not semantical, for it points to a difference in status and role not in meaning. Semantically speaking, to define A as B is to equate A with B. So much so that a justification of a definition may consist in proving the identity of definiens and definiendum: no such proof would be needed if every definition were just a linguistic convention.

The narrow construal of definitions as equalities has the following advantages. First, the identity in meaning (sense and reference) of definiendum and definiens is brought out. (On the other hand an equivalence does not warrant meaning identity since equivalents, though coextensive, need not be cointensive. More on this in a moment.) Second, every definition of the form $\ulcorner A =_{df} B \urcorner$, with A and B propositions, entails the equivalence $\ulcorner A$ iff $B\urcorner$ (or $\ulcorner$For all x, Ax iff $Bx\urcorner$) but not conversely. (If A and B are identical then either can take the place of "the other" in the tautology $\ulcorner A$ iff $A\urcorner$.) Third, every formally correct definition belongs by right in the class of analytic formulas – by Definition 1 in Sec. 2.1. Consequently the old dispute as to whether a convention might be true is settled: All definitional equalities are necessary truths, albeit rather cheap ones. (In fact they are the only necessary truths found in factual science apart from the formal truths employed in inference.) Fourth, any sentence expressing a definitional equality belongs to the object language of the conceptual system concerned: no further language level need be invoked since '$=_{df}$' is no longer a metalinguistic sign but just a metatheoretical status symbol. Fifth, proofs of concept independence may often be simplified, as identities are usually easier to check than equivalences. Moreover instead of employing semantic techniques (e.g. Padoa's) for checking concept independence (or undefinability) in a context one may try the following alternative procedure. Check whether definiens and definiendum meet the necessary condition of coextensiveness: if the extensions do not coincide then the suspected definiens will differ from the definiendum. However, if the *LHS* and the *RHS* are coextensive nothing can be concluded: the technique is one for disproving not for proving concept independence.

In the literature several alternative forms are used or even prescribed for definitions, chiefly the following:

$$\text{`}A\text{' names (designates) } B \qquad (D)$$
$$\text{``}A\text{'' means the same as ``}B\text{''} \qquad (S)$$
$$A \text{ iff } B. \qquad (E)$$

However, each of these formats has some disadvantage from which that of equality is free. In fact *D* multiplies unnecessarily the levels of language and does not guarantee substitutivity. *S* is not applicable prior to interpretation (Padoa, 1901) and moreover it presupposes either an intuitive (muddled) concept of meaning or a universally accepted theory of meaning – which, alas, is not forthcoming. Moreover if *S* is adopted then it prevents us from defining "meaning" without circularity. In any case, notwithstanding a popular view, a definition should not be required to "explain the meaning" of the definiendum – except of course in a pragmatic sense. Otherwise implicit definitions should be unacceptable. Consider, if in doubt, the recursive definition of the addition operation in the relational structure $\langle N, S \rangle$, viz.,

$$x+0=x \quad \text{and} \quad x+Sy=S(x+y) \qquad Df$$

where 'Sx' designates the successor of x. Since "+" appears on both sides, it is neither explained in terms of the previous concepts nor eliminated in favor of them (Goodstein, 1968).

Finally, the equivalence format *E* is open to the following objections. First, equivalents, though coextensive, are not necessarily cointensive – hence they may not be substituted for each other without a concomitant alteration in meaning. Second, it is important to keep the difference between equivalence and equality. Thus the biconditional $\ulcorner A$ is provable iff A is a tautology$\urcorner$ is a metalogical theorem not a definition of provability. Likewise Tarski's principle $\ulcorner$"s" is true iff $s\urcorner$ is a biconditional not a partial definition of truth.

Finally let us emphasize that only signs and their designata can be defined (provided they are not primitive). Factual items can be described, explained or predicted: not being constructs, facts cannot be built out of constructs. In sum there are no "real definitions". Hence, *pace* Suppes (1967, 1969) and Freudenthal (1970, 1971), the technique of axiomatic definition, ideal as it is for characterizing mathematical objects, cannot be expected to define the concrete objects described (not defined) by scientific theories (see Salt (1971) and Bunge (1973b)). For example, the equations of mechanics define (implicitly) a concept of body not bodies. So much for the misprint 'axiomagic'.

Like other methodological categories, that of definition has several dimensions: linguistically it is usually an abbreviation; epistemologically it is a way of building new concepts out of old; pragmatically it is a heuristic device and often a time saving device. We have dealt only with the semantic aspect of definition and cannot go any further.

2.3. *Presupposition*

A presupposition is a tacit or undisclosed supposition – one that can and should be brought to light by analysis. We shall distinguish six concepts of presupposition, all of them relevant to the analysis of scientific knowledge. They shall be introduced by so many definitions. In each of these 'A' and 'B' will designate sets of formulas.

First the *proof theoretic* notion of presupposition:

DEFINITION 10.2 *B presupposes A relative to the proof P* $=_{df}$ A belongs to the set of premises occurring in the proof P of B.

Change the proof and B may cease to presuppose A.

Next a first *semantic* concept of presupposition:

DEFINITION 10.3 *B weakly presupposes A meaning-wise* $=_{df}$ A is sufficient to determine the meaning of B.

For example, if a scientific theory B involves the concept of time then B must weakly presuppose, *inter alia*, some theory A of time. (The whole lot of theories presupposed meaning-wise by any given theory may be called the ***background*** of the theory: Bunge, 1967b.) Any number of alternative presuppositions may do the job, though not all of them equally well. Hence no single presupposition of this kind may be necessary.

For example, Newtonian mechanics can be axiomatized by presupposing either absolute space and time or relational space and time.

DEFINITION 10.4 *B strongly presupposes A meaning-wise* $=_{df}$ *A* is necessary to determine the meaning of *B*.

For example, arithmetic is a strong presupposition of any quantitative theory, both meaning-wise and proof theoretically.

Let us now distinguish two concepts of *alethic* presupposition:

DEFINITION 10.5 *B weakly presupposes A truth-wise* $=_{df}$ The truth of *A* is sufficient for the truth of *B*.

In other words, *B* is a weak alethic presupposition of *A* just in case $A \models B$. Consequently this concept of presupposition belongs to model theory. Therefore it has only an indirect relation to factual science, namely via mathematics. In factual science the most we can do is to check whether *B*, or rather some consequences of *B* conjoined with some extra premises (e.g. data), are approximately true. To put it otherwise: In factual science the most useful concept of entailment is the syntactic ($\vdash$) not the semantic ($\models$) one, since we must feel free to investigate the logical consequences of any assumption before assigning the latter any truth values. (Note that, although $\vdash$ and $\models$ are coextensive, they fail to be cointensive.) Next comes

DEFINITION 10.6 *B strongly presupposes A truth-wise* $=_{df}$ The truth of *A* is necessary and sufficient for the truth of *B*.

If weak alethic presupposition has a remote relation to factual science, strong alethic presupposition is even further removed. Finally we have the *methodological* concept of presupposition:

DEFINITION 10.7 *B presupposes A methodologically* $=_{df}$ The test of *B* uses *A* without questioning *A*.

For example, although electromagnetism is a fundamental theory in the sense that it can be formulated without resorting to any other scientific theory, its empirical tests presuppose a number of other theories – in fact as many as needed to design and control the instruments employed in the tests. Operationism could easily muddle up this issue by claiming that, this being so, then every theory presupposes every other theory. As a matter of fact the so called Copenhagen interpretation of quantum mechanics claims that, although quantum mechanics retrieves classical mechanics without being fully consistent with it, it also presupposes the latter because the former theory must be interpreted in terms of experiments, and every experiment must be accounted for by classical physics (Landau and Lifshitz, 1958, p. 3, criticized by Bunge, 1970b, pp. 310 ff.).

This winds up our examination of the (tiny) impact of our semantics on logical theory.

3. EPISTEMOLOGY

3.1. *The Status of Epistemology*

The collection of opinions concerning human knowledge, i.e., epistemology, used to be the nucleus of modern philosophy. Over the past century three other lines of inquiry have come to dispute the territory of epistemology: biology (following Helmholtz and Mach), psychology (following Piaget), and semantics (after Tarski and Carnap) – not to speak of the philosophy of science, which overlaps with epistemology, and pragmatics, which is still in the foetal state. Each of those disciplines offers persuasive arguments for absorbing the whole of epistemology. Let us weigh them.

The argument for surrendering epistemology to biology and psychology seems compelling: Perception and ideation are but so many aspects of man's effort to adapt himself to, as well as alter, his environment. Hence they belong to the study of man the animal. More particularly since sensing, perceiving, representing, and inferring are functions of the central nervous system, the study of knowledge falls within the province of neurophysiology and psychology.

All this is quite true. It has become ludicrous to speculate about perceiving without regard to the psychology of perception, about knowing without looking into learning theory, and so on. Some of the traditional problems of epistemology have indeed been snatched away by psychology. Epistemology is thus becoming biologized as demanded by Campbell (1959). This process is irreversible despite the efforts of philosophical psychology, which can now thrive only on the shortcomings of scientific psychology. Still, philosophy has a right to investigate knowledge (and for that matter anything) provided it does so from a different angle, with its distinctive means and aims, and as long as it learns from scientific psychology.

The philosophical study of knowledge includes a theoretical examination of the following subjects: (*a*) general nature of man's knowledge of his environment and of himself – i.e. knowledge as a metaphysical subject (in N. Hartmann's words, a subject of the *Metaphysik der Erkenntnis*); (*b*) kinds of knowledge (intuitive, rational, etc.) and their interrelationships; (*c*) relation between physical object statements and sense data statements; (*d*) concepts of factual truth, error, and error correction; (*e*) traditional dichotomies such as subjective-objective, a priori-a posteriori, experiential-conceptual, and intuitive-rational. None of these themes is claimed by biology or psychology – nor can they be studied seriously without the help of these disciplines.

As to the credentials of semantics, they too look impressive: while traditional epistemology dealt with truth in a metaphorical way and neglected meaning altogether, semantics offers exact theories of each of them. This is true, but still it may rejoined (*a*) that those exact theories are not concerned with factual knowledge, which is the main theme of epistemology, and (*b*) semantics pays no attention to the problems listed above. So, epistemology does have a territory of its own *vis-à-vis* semantics. Moreover, it could be argued that semantics, or at least the semantics of factual knowledge, is but *a part of epistemology* – namely that portion concerned with the reference, sense, and adequacy of human knowledge in general as different from the cognitive process.

Whatever view on the semantics-epistemology relations one may adopt, two things seem certain. The first is, that the two fields overlap. The second, that it matters little whether there is a border line (rather than a stripe) between the two as long as something be done about the problems themselves. Let us pick two representative problems of epistemology and see what our semantics has to say about them.

3.2. *Representation vs. Instrument and Picture*

One of the dichotomies tackled by epistemology is that of *a priori* vs. *a posteriori* knowledge. Because 'a priori' is usually taken to signify 'prior to experience' and because 'prior' is an ambiguous word, 'a priori knowledge' can be a misleading expression. Thus while it is true that logic *holds* irrespective of experience, it is false that it *originated* independently of experience: logical investigation is just a fragment of human experience. For this reason – i.e., to avoid the genetic fallacy – it is advisable to draw a clear distinction between origin and validity, and to relativize the notion of *a priori* to a body of knowledge. The two requirements are met by the following conventions.

DEFINITION 10.8 The set S of statements is *a priori* relative to the body of knowledge $K =_{df}$ No member of S presupposes any member of K either meaning-wise or truth-wise.

DEFINITION 10.9 The set S of statements is ***absolutely a priori*** $=_{df}$ S is *a priori* relative to every body of knowledge.

In this sense logic is absolutely a priori, mathematics is a posteriori (=not a priori) with respect to logic, certain branches of mathematics are a posteriori relative to others, science is a posteriori relative to the whole of formal science, and certain branches of science are a priori relative to others. These epistemological notions of a priori and a posteriori are not the same as the psychological and the methodological concepts similar to them. For example, a new conjecture in logic is a priori and moreover absolutely so, but it is methodologically a posteriori in the sense that it must be checked before being included in logic. And a new scientific hypothesis, though a posteriori in the epistemological (or semantical) sense, is psychologically a priori in the sense that it comes to us before any new data.

Whatever bit of human knowledge is prior to some other bit may be said to function as an *instrument* for the latter. The epistemological questions are (*a*) whether the whole of factual science is but an instrument for action – as pragmatism and the anti-science movement hold, and (*b*) whether every scientific theory is nothing but a data processing instrument (conventionalism, nominalism, pragmatism, and computerism). That factual knowledge can be used as an instrument, for good or for evil, is beyond dispute: the question is whether, in addition, it represents reality (or at least experience) and if so how.

There are, of course, as many answers to the above question as there are epistemological schools. The most popular are direct (naive) realism and classical empiricism, according to which sensations and ideas, even if logical or mathematical, represent and moreover reflect or copy factual items. According to critical realism, on the other hand, our perceptual and conceptual representations of external objects are "only *signs* [symptoms] of external objects, and in no sense *images* of any degree of resemblance. An image must, in certain respects, be *analogous* to the original object (...) For a *sign* it is sufficient that it become apparent as often as the occurrence to be depicted makes its appearance, the conformity between them being restricted to their presenting themselves simultaneously" (Helmholtz, 1873, p. 391). This is, roughly, the view we adopted in Chs. 2 and 3: factual knowledge *refers* to external objects and *represents* them but it is *symbolic* rather than pictorial. Moreover the representation it effects is one-sided rather than full, and *global* (whole theory-factual domain) rather than pointwise. (Cf. Ch. 6, Sec. 3.4.) How else are we to explain that scientific theories so often go wrong and that they do not depict anything although they explain and predict?

Theoretical concepts and scientific theories do not *picture* physical objects, let alone pointwise. (Cf. Ch. 3, Sec. 1.) They cannot picture because they are constructs. Of course they may represent and, as long as they are factual, they all refer to supposedly real items. However, only some of the elements of a theory represent: others, like lagrangians and partition functions, discharge no representational function even if they do refer. (Recall Ch. 3, Sec. 1.1, and Ch. 7, Sec. 3.4.) Conversely there are real traits, such as individual idiosyncrasies, that no scientific theory will capture unless it happens to be a theory of an individual. Even so, no scientific theory attempts to provide a complete description of its referents. (Consequently Bohr and Heisenberg were wrong in claiming that quantum mechanics provides a complete description, and Einstein was right in holding that it does not. But Einstein was misguided in his search for an alternative theory that would supply "a complete description of reality". It is not only that a description requires empirical data in addition to theories if it is to describe individual things: both the history of science and an analysis of the way theories are constructed suggest that the process of taking in more and more factual items is never ending, as stressed by dialectical materialists: see Lenin (1909).)

Furthermore theoretical concepts and theories are not abstracted from sense experience, not even from scientific experiments, so they should not be placed in the same bag with empirical concepts. For one thing they need not have the same referents. For another they never have the same sense: for, if they were cointensive, then the theory concerned would be pointless. Of course there is some correspondence between concepts and percepts, but (*a*) it is not one to one and (*b*) it is symbolic or indirect (Einstein, 1936, p. 353). Materialists need not panic: the view that our representations of the external world are symbolic rather than cinematographic is perfectly consistent with the theses that there are no autonomous ideas and that our ideas, if they happen to concern autonomous external objects, sometimes do succeed in representing the latter in a more or less adequate way. In any event the point we wished to make is that critical realism is more adequate than the picture view of knowledge, which has not gone beyond the metaphor stage anyway. Nonetheless it must be owned that the symbol doctrine, proposed one century ago, is in dire need of elaboration. We are still far away from the systematic realism demanded by Hooker (1974).

3.3. *Objectivity vs. Subjectivity*

A representation, by definition, represents something: there can be no representation in and by itself, or *Vorstellung an sich.* In particular, a conceptual representation of a factual item x is a construct y that maps (some of) the features of the object x. A representation is not a part, let alone the whole, of its object. So much for the semantics of the concept of representation – some of the complexities of which were examined in Ch. 3. In epistemological terms: "Unless the content of knowledge is recognized to have a condition independent of the mind, the peculiar significance of knowledge is likely to be lost. For the purpose of knowledge is to be true to something which is beyond it" (Lewis, 1929, p. 192). This is, of course, the realist or objectivist thesis.

The subjectivist, on the other hand, holds that the world is our representation or part of the latter. Thus Goodman claims that there is no such thing as *the* way the world is: the world is as many ways as it can be truly described, seen, pictured, etc. (Goodman, 1960). Further: "That nature imitates art is too timid a dictum. Nature is a product of art and discourse" (Goodman, 1968, p. 33). The rationale is that representations are creations not copies. True enough, but the point is that a representation, however creative, does not create its object. To claim that it does is to muddle things up (as Frege (1894) noted) and to make a mockery of the age long striving for objectivity, that supreme accomplishment of the creative scientist. For it is a declared aim of scientific research to obtain objective (impersonal and publicly testable) representations of the world. This is why scientists keep checking them and trying to improve on them. This is why any subjective elements, when spotted, are weeded out. And because of this need for keeping the distinction between objectivity and subjectivity, it behooves the semanticist to clarify it.

Whether or not a representation is objective is a matter for semantics and methodology, since a construct is said to be semantically objective if it concerns external objects alone and can be subjected to impersonal tests. (But of course those objects may be persons and they may even prove to be nonexistent.) On the other hand a construct *a* is a *subjective representation* of an object *b* just in case *a* refers not only to *b* but also to the subject *c* responsible for *a*. If this subject dependence is genuine then it must show up upon substituting a different subject for the original one. This is how it can be shown that quantum mechanics, reputedly a subjectivistic theory, is perfectly objective – namely by analyzing the referents of the basic predicates of the theory and proving the cognitive subject to be absent from them (Bunge, 1967e, 1973b).

A methodological examination should confirm or infirm the results of a semantical test for the objectivity of a representation. That examination may consist either in the control of the suspected variables or in some other empirical procedure. Consider a sequence of events of a kind assumed to be mutually independent. On a subjectivist interpretation the events should occur only when observations are actually performed – e.g. every minute. Consequently the distribution in time will be binomial or similar. On an objectivist interpretation the events will occur whether or not they are observed. In particular it may happen that the process goes on continuously over a time interval. Hence the distribution could be Poisson's or similar. We have then two rival distributions of elements among the various accessible states, i.e. two different conceptual representations or models. The conflict can be resolved by sampling. Needless to say, the problem hardly arises in the physical sciences. But it does arise in psychology, and in any case it is philosophically important to know that the problem of objectivity can in principle be solved by theoretical modeling and subsequent empirical tests.

We take it then that science espouses, albeit tacitly, a realist epistemology. Moreover science confirms critical (or indirect) realism rather than naive (or direct) realism, as shown by its distrust of the "immediately given" and its relentless effort to improve on every theoretical representation of facts. In order to better appreciable this point we list the main theses of these two versions of realism. (Cf. Bunge, 1973a.)

Direct (naive) realism	*Indirect (critical) realism*
$\mathcal{DR}$ 1 Everything is knowable.	$\mathcal{CR}$ 1 Many things and facts are knowable. Many others are not – e.g., extinct things that have left no perceptible traces.
$\mathcal{DR}$ 2 Physical objects are perceived directly: the eye is innocent.	$\mathcal{CR}$ 2 Perception can be improved or distorted by preconception (superstition, hypothesis, etc.): the eye is biased – for better or for worse.
$\mathcal{DR}$ 3 All objects are conceived directly.	$\mathcal{CR}$ 3 Constructs are formed within a body of background knowledge, unevenly true and largely social.
$\mathcal{DR}$ 4 Every conceptual representation is at least homomorphic.	$\mathcal{CR}$ 4 Conceptual representations are rarely pointlike (e.g. homomorphic), mostly global (whole theory-whole factual domain).
$\mathcal{DR}$ 5 All sensations and conceptions mirror real objects.	$\mathcal{CR}$ 5 Some constructs are representational while others are purely syntactic. And many hypotheses and theories intended to represent real factual items turn out to be wholly fictitious.
$\mathcal{DR}$ 6 Perfect knowledge (complete and fully true) is possible.	$\mathcal{CR}$ 6 Factual knowledge is always imperfect (incomplete and at best partially true). But it is perfectible.

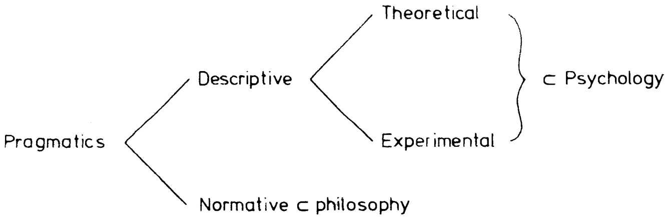
Note the following points. First, although realism (whether direct or indirect) presupposes the ontological hypothesis that there are things in themselves, i.e., objects existing independently of any mind, it does not espouse materialism. Indeed a realist need not entertain the ontological hypothesis that every existent is material. (Nor does materialism entail realism: a materialist need not believe that matter can be known: witness Spencer.) Second, realism does not entail rationalism: the former is consistent with Meyerson's thesis about the unavoidable and irreducible irrational remainder left by every cognitive venture (Meyerson, 1908, pp. 272ff). Third, realism is not committed to the causal theory of perception, i.e., the hypothesis that sense impressions are causally related to physical objects. It is possible to hold to realism while assuming that perceptions, though caused by external (and internal) objects, are stochastically (rather than causally) related to them. In any event it is for science, not for epistemology, to find out what the mechanisms of perception and ideation are. Epistemology is concerned with knowledge (the end product) rather than with cognition (the process). Which brings us to our next theme.

3.4. *The Knowing Subject*

There is no knowledge without both an object of knowledge and a knowing subject. The claim that there is absolute knowledge, or knowledge in itself, above and beyond concrete knowing subjects, is fantastic. Moreover it violates the very syntax of 'to know', for 'x is known' is short for 'There is at least one y such that y is a knowing subject and y knows x'. Obliterate mankind and no human knowledge will remain. Furthermore every individual learns, imagines, and remembers (in sum, knows) in his own way: real cognition is as personal as ignorance. For this reason knowing or cognition is a subject for psychology. Epistemology takes knowing for granted and focuses on what is alleged to be known. In other words, epistemology is not concerned with personal knowledge, the only cognition there is. Epistemology makes the useful pretense that there exists *impersonal knowledge*, just as semantics pretends that there are propositions – not just judgments and sentences – and just as mathematics pretends that there are proofs whether or not they have actually been produced by anyone. To call what is known, i.e., knowledge, a *world* and assume that it is superimposed on the world of fact (Popper, 1968) is an unnecessary Platonic fantasy. There is only one world and cognitive subjects are part of it and intent on knowing (or ignoring) some chunks of it. Human action, whether cognitive or of another kind, does not create new worlds (except metaphorically) but variously transforms the only world there is.

Some philosophers, following certain suggestions of Peirce's and Morris', hold that personal knowledge is to be studied by a special discipline concerned also with other facets of human activity. This discipline, *pragmatics*, would consist of an empirical branch and a philosophical one. Philosophical (or pure) pragmatics would seek to establish logical relations among pragmatic concepts such as those of belief and intention (Martin, 1959) or to develop theories such as tense logic and the logic of personal pronouns, in which the subject or user plays a central role (Montague, 1968, 1970).

It may be questioned whether pure pragmatics, as currently developed, is on a sound methodological basis. Indeed if a discipline is concerned with factual items, such as users and circumstances of use, without benefiting from any empirical investigations, then it belongs with speculative metaphysics – with *Naturphilosophie* or with *Kulturphilosophie.* Theoretical pragmatics, if descriptive of fact, should be methodologically similar to theoretical psychology: it ought to build mathematical models of certain aspects of human behavior and have them tested in the laboratory. Only when normative might pragmatics be exempted from this requirement. For example, although actual problem solving is studied by psychology, the analysis of *well* conceived and *well* posed conceptual problems in general can be the concern of philosophers. That is, the logic and semantics of problems (like the logic and semantics of norms) belong to philosophical pragmatics. In conclusion, we have the following branching:



The relation between philosophical pragmatics and semantics could be this: each should investigate its own side of the coin of knowledge, but pragmatics would, in addition, take up problems concerning action. To put it differently: Epistemology can be split into two branches, according as the knowing subject is or is not reckoned with explicitly: (*a*) the study of personal knowledge, which is a branch of pragmatics, and (*b*) the study of impersonal knowledge, which overlaps with semantics. For example, the very concept of knowledge will split up into those of a set of statements, and opinion (or belief); the concept of truth into those of objective truth and personal truth; and the concept of meaning into those of semantic meaning and pragmatic meaning (or normal linguistic usage in a community).

It is well known that pragmatics is not married to pragmatism: that one may hold either a pragmatist view of pragmatics or an alternative conception of it. The extreme pragmatist theses about pragmatics seem to be these: (*a*) All genuine syntactical and semantical concepts have pragmatic counterparts, and (*b*) the former should all be reducible to (e.g., definable in terms of) pragmatic terms. While the first thesis is interesting, possibly true, and therefore deserves being investigated, the reduction thesis (*b*) is false. For example, Wittgenstein's reduction of meaning to use is, to put it charitably, a proposal to redefine 'meaning' so as to suit the purposes of the lexicographer. Usage is nothing but an (unreliable) indicator of significance. Besides, there is such a thing as correct (as distinct from widespread) use. And the relation of this normative concept to the semantic concept of significance is roughly this: The correct use of a sign is determined by its significance as revealed by an analysis of the system in which it occurs.

Moreover the pragmatic concepts presuppose syntactic and semantic concepts not the other way around. For example, the pragmatic statements $\ulcorner t$ has been proved from $A\urcorner$ and $\ulcorner x$ claims to have proved t from $A\urcorner$ presuppose the metamathematical concept of proof. Without the latter the previous statements would be unintelligible and untestable. Likewise and for the same reason the pragmatic proposition $\ulcorner$'*a*' means "*b*" to $c\urcorner$ presupposes a semantic concept of meaning. The same holds for belief assertions: In order to believe p we must first know p, which in turn presupposes p (quite apart from any truth value assignment). On the other hand we can pretend that there are propositions known to nobody. (This is what we do when we play around with the totality $\mathscr{C}n(A)$ of consequences of a given set A of assumptions.) And, of course, we can know p without believing p, just as we can believe p without professing to understand p.

A final warning: by an egocentric habit acquired in infancy we tend to employ pragmatic terms where they do not belong, thus contributing to an inflation of pragmatics. For example, we are prone to say that organic compounds are *found* in certain stars instead of asserting that *there are* organic compounds in stars; or that p is a possible result of *measuring* P rather than stating that p is a possible value of P; or that t

can be proved from A, instead of saying that t follows from A. The fact of the matter is that x may be found (or measured) if it *exists* to begin with, and that t may be proved provided it *is* a consequence. The phony use of pragmatic terms only creates confusion and fosters the illusion that pragmatics is a going concern rather than a project. In any case the semanticist's mode of speech is objective (subject-free) and tenseless: he considers matters *sub specie aeternitatis.*

So much for the relation of semantics to epistemology. Let us finally take a look at the most substantial yet also least reputable neighbor of semantics – metaphysics.

4. METAPHYSICS

4.1. *The Metaphysical Neutrality of Language*

We met metaphysics at the outset (Ch. 1, Sec. 3) when listing the kinds of object we should distinguish, as well as when discussing the nature of constructs. We also encountered metaphysics along the way, particularly when discussing reference, the representational function of some scientific constructs and, tangentially, factual truth. There are several other contacts between semantics and metaphysics. And, if the two are approached with a modicum of reason, their contacts need not be points of friction. Semantics can supply metaphysics with some tools for dissipating conceptual obscurity and confusion, and metaphysics can reciprocate by helping semanticists to avoid the extremes of vulgar materialism (e.g., literalism) and Platonism. We shall not pursue this line here. Instead, we will point to a place or two where semantics does *not* meet metaphysics. In fact we propose to discuss the alleged ontological commitment of language (the Whorf-Sapir doctrine) and, in particular, the claim that all existence sentences make such a commitment (Quine's thesis).

The celebrated Whorf-Sapir hypotheses are, in a nutshell, that (*a*) ordinary languages are loaded with world views or metaphysics, and (*b*) one's language determines at least in part the way he perceives and conceives the world. These two conjectures are, to say the least, gross exaggerations of the undeniable feedback of language into thought. It is false that every language reflects some world view, let alone some metaphysical system: any advanced language, by definition of 'advanced', can express any number of mutually incompatible world views, as witnessed by the diversity of metaphysics expounded in Sanskrit or in ancient Greek. And it is false that language is an important causal factor in perception and conception: a language permits or restricts our expressing what we perceive or conceive rather than what we perceive or conceive. Perception and conception occur in an epistemic framework not in a linguistic one: what determines in part whatever we perceive and conceive is experience, whether spontaneous or disciplined, personal or collective, rather than our mother tongue. Language is fortunately neutral with respect to our sensing and conceiving the world. Otherwise it would be impossible to formulate and discuss mutually incompatible views, and the very concept of expressive power of a language would be devoid of meaning.

Moreover the Whorf-Sapir doctrine is refuted by educational psychology: verbal teaching is inefficient unless the subject has already acquired some of the ideas he has been talked to about. If no words are available to express new ideas then one invents them at any age and in any culture. But if there are no ideas then words won't help – except to disguise the poverty in ideas, as Mephisto pointed out. Piaget and his coworkers have shown that a child who has learned the correct use of 'long', 'short' and cognate words may fail to understand that the amount of clay does not change when turned under his eyes from a ball into a sausage. Here again what matters is knowledge not language. In sum the Whorf-Sapir hypotheses are false: language is a tool that carries no ontological burden. It is rather the other way around: every ontology is constrained by the language it employs, in the sense that the latter may or may not be rich enough to express certain ontological ideas. (For example, ordinary language is insufficient to formulate a general theory of space.) Only theories can be ontically committed, and they are so provided they are not theories in logic or in pure mathematics. Language cannot even suggest a bad theory. But we went through this in Ch. 1: let us turn to a more interesting problem.

4.2. *The Metaphysical Neutrality of Logic*

If logic were nothing but a language then the arguments of the previous subsection would apply to it: logic should be ontologically neutral because it is the universal language. But logic is also and primarily a theory

– the theory of deduction – even though it can be used as a language (recall Ch. 1, Sec. 1.3). Therefore the relation of logic to metaphysics must be investigated separately. However, this is not the place for a thorough investigation of this problem, treated elsewhere (Bunge, 1974a). We must restrict our attention to those aspects of the question that involve semantics. We will limit ourselves to the question whether logic, in addition to being a reasoning organon, has a semantics similar to that of factual science. In other words, we shall ask the question whether logic involves a reference to reality or is ontologically neutral.

The thesis that logic does call for an ontological interpretation has been championed in recent times by Scholz and by Quine though on different grounds. According to Scholz (1941) the logical formulas hold for all possible worlds, ergo in particular for ours: logic would thus constitute a minimal ontology. This argument hinges on the identification of "possible world" and "model". But this is just a verbal trick: whereas a model is a conceptual object, the actual world is not a model of an abstract structure but the maximal self existent thing, and a possible world is something that neither logic nor ontology are in a position to characterize. Logic does not hold in all possible worlds – except metaphorically, in the sense that it is independent of the constitution and structure of the world. What model theory says is something more modest, namely, that a formula is a logical formula just in case it is satisfied in all models, i.e., if it holds under all interpretations of the nonlogical variables it contains. And the model independence of a logical identity consists in a particular matching of it with a whole set of constructs: it is an intraconceptual affair that has nothing to do with the actual world (Ch. 6, Sec. 2). Therefore the notion of universal validity (or analyticity) does not involve that of factual truth, which characterizes the semantics of factual science (Ch. 8). In conclusion, model theory is not concerned with any worlds and it does not exhibit logic as an ontology.

Unlike Scholz' approach, Quine's is direct: the latter finds ontology in the very bones of logic, particularly in quantification. Indeed he holds that "referential quantification is the key idiom of ontology" (Quine, 1969, p. 66). More explicitly: "Existence is what existential quantification expresses. There are things of kind F if and only if $(\exists x)\, Fx$. This is as unhelpful as it is undebatable, since it is how one explains the symbolic notation of quantification to begin with" (*op. cit.*, p. 97). The set of "things" a theory takes to exist is called by Quine the "ontological commitment", at times also the "ontology", of the theory (*op. cit.*, p. 106). Actually the word 'commitment' is too strong, since "that commitment can be momentary" (Quine, 1970, p. 99). That is, we have, if at all, ontological hypotheses.

If 'commitment' is unfortunate so are 'ontological' and 'ontic' in this context. Indeed, all that is intended to point to is the collection of *referents*, whether hypothesized or certified, intended or unintended, physical or conceptual, of a theory couched in some system of predicate logic. For example the referents of arithmetic are integers or perhaps the field of rationals. But a statement such as ⌜There are infinitely many prime numbers⌝ commits no one to believing in the *autonomous* existence of prime numbers. Far from being "an ontic idiom *par excellence*" (Quine, 1970, p. 92) existential quantification, unless qualified, is ontologically neutral. It is up to factual science, not to philosophy, to decide whether an unqualified existence statement has an ontic import.

The existential quantifier can be eliminated in favor of negation and the universal quantifier: "There are F's" is short for "It is not the case that everything fails to be an F". On the other hand, the concept of physical (real, ontic) existence cannot be so eliminated. In fact, to indicate physical existence we must assert unqualified existence and add that the objects in question happen to be physical objects – where the predicate 'is a physical object' is elucidated in ontology not in logic. There are *real* things of kind F if and only if $(\exists x)$ (Fx & x is a physical object). Only objects of this kind are of possible interest to ontology conceived, not as the general theory of objects of any kind, but as general cosmology. And only such statements of physical (real, ontic) existence will "commit" us to the objects they refer to – or rather to the task of finding out whether such objects are part of the actual world. (Incidentally, the collection of such referents of a statement constitutes its reference class, not its ontology: an ontology is a theory.)

Conceptual existence is parallel. There are *conceptual* objects of kind F if and only if $(\exists x)$ (Fx & x is a construct). This is the kind of existence the mathematician is interested in. Thus, notwithstanding the nominalist prohibition against classes, the mathematician deals with them all the time – only, he does not claim more than conceptual existence for them.

He has no reasons for thinking that the members of a set in a mathematical theory are more real than the set itself: if assumed or proved to be nonempty, a class enjoys the same conceptual existence as its members. Surely the metaphysician is entitled to asserting that classes have no physical existence – but then the mathematical individuals are just as unreal as classes, so that the metaphysician's remark is irrelevant.

In sum, while logic needs only a general concept of existence, the moment we cross the border of logic we need two specific concepts of existence – conceptual and physical. We have thus at least three different concepts of existence: neutral, conceptual, and physical. (These are concepts of existence not modes of being. Consequently there is no question of a general ontology branching off into an ontology of things and another of constructs. Let us not get carried away by words.) Let us emphasize the differences:

$(\exists x)\, Fx$	*Neutral existence*
$(\exists x)\,(Gx \,\&\, x$ is a construct$)$	*Conceptual existence*
$(\exists x)\,(Hx \,\&\, x$ is a physical object$)$	*Physical existence.*

Correspondingly we have three kinds of existence problem: neutral, conceptual, and physical. The last two kinds are irreducible. Thus the question of the (conceptual) existence of classes has nothing to do with the question of the (physical) existence of things: this is why conceptualism is compatible with ontological individualism, or the belief that only individuals can have physical existence. And, since all classes are constructs, infinite sets pose no special problem of existence.

Nor are those kinds of existence question reducible to a fourth, more basic type of question. In particular it is not possible to reduce all questions of existence to the existence *in* or *of* a linguistic framework – the so called internal and external questions (Carnap, 1950). Thus the problem whether a "theoretical entity", such as the parton or the black hole, has a real counterpart is not a question concerning exclusively a man made "framework", let alone a linguistic one: it is an empirical problem – namely that of subjecting the factual theory concerned to observational tests. To hold that physical existence questions are not scientific but concern linguistic frameworks is no better than claiming, with Mach and the Vienna Circle, that those are sheer metaphysical questions. In either case the typical features of the semantics of scientific theories are missed. (For further criticisms see Ferrater Mora (1967).)

The upshot of our discussion is this: (*a*) just as we distinguish constructs from things so we must distinguish conceptual existence from physical existence; (*b*) logic employs a generic or neutral concept of existence, hence it is not in a position to assert or deny anything concerning the real world: it is ontologically neutral; (*c*) mathematics uses the more special notion of conceptual existence but it is equally unconcerned with reality; (*d*) factual science and metaphysics utilize the concept of physical (material) existence.

4.3. *Metaphysical Commitments of the Semantics of Science*

Logic, mathematics, and their semantics (i.e., model theory) are metaphysically neutral: they do not even have to assume the existence of any objects other than constructs and their symbols. Not so the semantics of factual science. Indeed the latter has the following presuppositions of a metaphysical character: (i) there are both constructs (in particular concepts) and physical objects (in particular signs); (ii) some signs designate constructs and some constructs refer to physical objects; (iii) the theories referring to physical objects constitute more or less adequate (true) representations of aspects of the world.

If the above assumptions were denied then the very concepts of designation, denotation, reference, representation, and factual truth would become pointless. Our own realist brand of semantics adds two further assumptions: (iv) constructs are ideal objects: unlike physical objects, constructs have no separate (physical) existence, and (v) the putative referents of a factual construct are supposed (rightly or wrongly) to be self existent not just possibilities of perception.

This is all the metaphysical load our semantics carries. It is necessary and sufficient to make sense of the rules of denotation and of the semantic assumptions included in scientific theories (when reconstructed according to our semantics) as well as to make sense of the testing for truth of fact and of the correction of hypotheses and theories in the light of fresh experience. Beyond this point there is room for any number of possible ontological theories sketching the basic constitution and structure of the world. The author's own preference lies with cross disciplinary metaphysical theories, mathematical in form and contiguous with factual science (Bunge, 1971b). But this is another story to be told in a different work. (Cf. Volumes 3 and 4 of this *Treatise.*)

5. PARTING WORDS

Up till now philosophical semantics has led a rather secluded life: it has had hardly any contacts with other branches of philosophy and with factual science. Whence its main failing: its nearly total irrelevance to anything lying outside logic and mathematics – and consequently its inability to be of any help in understanding what goes on in that external world.

It has become an urgent task to reset semantics in its philosophic context – the one in which it used to dwell from Socrates through Buridan to Russell – as well as to open it up to the winds of factual science, without thereby relinquishing the ideals of rigor and systemicity bequeathed upon us by Tarski and Carnap.

Secondly, semantics should become a service discipline like logic and mathematics – one ever ready to offer a helping hand to whichever neighbor may need it. (See Table 10.1.) But in order to become useful semantics must mix with its neighbors and even meddle with their affairs, thus learning their ways and needs. The watchword should no longer be 'Come and behold my neatness' but rather 'Let us go and see how we can fix that mess'.

Thirdly, semantics should be freed from its associations with obsolete philosophies, and fresh semantic theories answering the needs of science and inspired in critical realism should be built. We need far more work in realist semantics and we should look everywhere for inspiration and control: a philosophy that minds only its own business does not mind its main business.

May this work contribute to implementing Kanenas T. Pota's ideal of "an exact, systematic, realist, and above all useful semantics". In any case this semantics of ours is to serve as a prolegomenon to the substantive portions of the philosophical system set forth in this *Treatise*, namely ontology, epistemology, and axiology.

TABLE 10.1

Semantics and its neighbors

Trait	Science	Mathematics	Philosophy			
			Logic	Semantics	Epistemology	Metaphysics
Referents	Things (concrete systems)	Conceptual structures (categories)	Constructs	Constructs & signs	Knowledge	Reality
Range	Aspects and levels of things	Theories	Universal	Universal	Universal	Universal
Truth type	Factual	Formal	Logical	Philosophical	Philosophical	Philosophical
Accuracy	Limited	Complete (except in numerical analysis)	Complete	Complete with reference to formal science	Limited	Limited
Method	Hypothesis & theory, observation & experiment	Postulation & proof, example & counter-example	Postulation & proof, example & counter-example	Postulation, proof, & checking with logic, mathematics, or science	Analysis, postulation, & checking with substantive knowledge & methodology	Postulation & checking with science
Goals	Finding laws, describing, explaning & predicting	Building theories about conceptual structures & inter-relating them	Depuration & systematization of general methods of analysis and proof	Elucidation & articulation of concepts of meaning, truth & cognates	Elucidation & articulation of all the concepts about factual knowledge & ignorance	Finding overall (cross-disciplinary) structures & patterns in the world

Table 10.1 (Continued)

Trait	Science	Mathematics	Philosophy			
			Logic	Semantics	Epistemology	Metaphysics
Roles	Basing technology, controlling world views	Forging conceptual tools for science & philosophy	Keeping watch on reasoning	Conceptual hygiene, spotting genuine referents, clarifying sense, and scotching myths in the philosophy of scientists	Methodological alertness & open-mindedness	Elucidating concepts common to several sciences, posing fruitful problems & proposing illuminating hypotheses about the world
Employs	Logic, mathematics & philosophy	Logic	Mathematics	Logic, mathematics & science	Logic, mathematics & science	Logic, mathematics & science

BIBLIOGRAPHY

Addison, J. W., L. Henkin, and A. Tarski, Eds. (1965), *The Theory of Models*. Amsterdam: North-Holland Publ. Co.

Ajdukiewicz, K. (1967a). Intensional expressions. *Studia Logica* **20**: 63–86.

Ajdukiewicz, K. (1967b). Proposition as the connotation of sentence. *Studia Logica* **20**: 87–98.

Alston, W. P. (1968). Meaning and use. In Parkinson Ed. 1968: 141–165.

Arnauld, A. and P. Nicole (1662). *La logique, ou l'art de penser*. Paris: Flammarion 1970.

Attfield, R. and M. Durrant (1973). The irreducibility of 'meaning'. *Nous* **VII**: 282–298.

Austin, J. L. (1962). *How to do Things with Words*. London: Oxford University Press.

Ayer, J. A. (1959). *Logical Positivism*. Glencoe, Ill.: Free Press.

Bar-Hillel, Y. (1970). *Aspects of Language*. Jerusalem: Magnes Press.

Bar-Hillel, Y., E. I. J. Poznanski, M. O. Rabin, and A. Robinson, Eds. (1961). *Essays on the Foundations of Mathematics*. Jerusalem: Magnes Press.

Bar-Hillel, Y., Ed. (1965). *Logic, Methodology and Philosophy of Science*. Amsterdam: North-Holland Publ. Co.

Barcan-Marcus, R. (1962). Interpreting quantification. *Inquiry* **5**: 252–259.

Bell, J. L. and A. B. Slomson (1969). *Models and Ultraproducts*. Amsterdam: North-Holland.

Benacerraf, P. (1973). Mathematical truth. *Journal of Philosophy* **70**: 661–679.

Beth, E. W. (1962). Extension and intension. In *Logic and Language: Studies dedicated to Professor Rudolf Carnap*. Dordrecht: D. Reidel Publ. Co.

Birkhoff, G. (1961). *Lattice Theory*, 2nd ed. Providence, R. I.: American Mathematical Society.

Black, M. (1962). *Models and Metaphors*. Ithaca: Cornell University Press.

Blanshard, B. (1939). *The Nature of Thought*. London: Allen & Unwin; N. York: Mcmillan.

Blumenthal, L. M. and K. Menger (1970). *Studies in Geometry*. San Francisco: W. H. Freeman and Co.

Bolzano, B. (1837). *Wissenschaftslehre*, 4 vols. Sulzbach: Seidelsche Buchhandlung. Abridged Engl. transl. by R. George, *Theory of Science*. Berkeley and Los Angeles: University of California Press 1972.

Bolzano, B. (1851). *Paradoxes of the Infinite*. Transl. D. A, Steele. London: Routledge & Kegan Paul, 1950.

Bunge, M. (1957). Lagrangian formulation and mechanical interpretation. *Amer. J. Phys.* **25**: 211–218.

Bunge, M. (1959a). *Causality*, Cambridge, Mass.: Harvard University Press. Repr.: The World Publ. Co. 1963.

Bunge, M. (1959b). *Metascientific Queries*. Springfield, Ill.: Charles C. Thomas Publ.

Bunge, M. (1961a). Laws of physical laws. *Amer. J. Phys.* **29**: 518–529. Repr. in Bunge (1963a).

Bunge, M. (1961b). Analyticity redefined. *Mind* **70**: 239–246.

Bunge, M. (1961c). The weight of simplicity. *Phil. Sci.* **28**: 120–149.

Bunge, M. (1962). *Intuition and Science.* Englewood Cliffs, N. J.: Prentice Hall.

Bunge, M. (1963a). *The Myth of Simplicity.* Englewood Cliffs, N.J.: Prentice-Hall.

Bunge, M. (1963b). A general black box theory. *Phil. Sci.* **30**: 346–358.

Bunge, M. (1964). Phenomenological theories. In M. Bunge, Ed., *The Critical Approach,* pp. 234–254. New York: Free Press.

Bunge, M. (1966). Mach's critique of Newtonian mechanics. *Amer. J. Phys.* **34**: 585–596.

Bunge, M. (1967a). *Scientific Research,* 2 vols. New York: Springer-Verlag.

Bunge, M. (1967b). *Foundations of Physics.* New York: Springer-Verlag.

Bunge, M. (1967c). Analogy in quantum mechanics: from insight to nonsense. *Brit. J. Phil. Sci.* **18**: 265–286.

Bunge, M. (1967d). Physical axiomatics. *Rev. Mod. Phys.* **39**: 463–474.

Bunge, M. (1967e). A ghost-free axiomatization of quantum mechanics. In M. Bunge, Ed., *Quantum Theory and Reality,* pp. 98–117. New York: Springer-Verlag.

Bunge, M. (1969). What are physical theories about? *Amer. Philos. Quart. Monograph* **3**: 61–99.

Bunge, M. (1970a). Theory meets experience. In H. Kiefer and M. Munitz, Eds., *Contemporary Philosophic Thought,* **2**: 138–165. Albany, N.Y.: State University of New York Press.

Bunge, M. (1970b). Problems concerning inter-theory relations. In P. Weingartner and G. Zecha, Eds., *Induction, Physics, and Ethics,* pp. 287–325. Dordrecht: D. Reidel Publ. Co.

Bunge, M. (1971a). A mathematical theory of the dimensions and units of physical quantities. In M. Bunge, Ed., *Problems in the Foundations of Physics,* pp. 1–16. New York: Springer-Verlag.

Bunge, M. (1971b). Is scientific metaphysics possible? *J. Phil.* **68**: 507–520.

Bunge, M. (1971c). A new look at definite descriptions. *Phil. Sci. (Japan)* **4**: 131–146.

Bunge, M. (1972a). A program for the semantics of science. *J. Phil. Logic* **1**: 317–328. Reprinted in Bunge, Ed. (1973a).

Bunge, M. (1972b). Metatheory. In *Scientific Thought.* Paris–The Hague: Mouton/Unesco.

Bunge, M. (1973a). *Method, Model and Matter.* Dordrecht: D. Reidel Publ. Co.

Bunge, M. (1973b). *Philosophy of Physics.* Dordrecht: D. Reidel Publ. Co.

Bunge, M. (1973d). Meaning in science. *Proc. XVth World Congress of Philosophy* **2**: 281–286. Sofia.

Bunge, M., Ed. (1973a). *Exact Philosophy: Problems, Goals, and Methods.* Dordrecht: D. Reidel Publ. Co.

Bunge, M. (1974a). The relations of logic and semantics to ontology. *J. Phil. Logic* **3**: 195–210.

Bunge, M. (1974b). Possibility and probability. In W. L. Harper and C. A. Hooker, Eds., *Proc. of an Internat. Research Colloquium on Foundations of Probability and Statistics,* Vol. 3. Dordrecht and Boston: D. Reidel Publ. Co.

Burton, W. K. (1968). Constructive thermodynamics. In A. A. Schmidt, K. Schütte, and H.-J. Thiele, Eds., *Contributions to Mathematical Logic,* pp. 75–89. Amsterdam: North-Holland.

Campbell, D. T. (1959). Methodological suggestions from a comparative psychology of knowledge processes. *Inquiry* **2**: 152–82.

Carathéodory, C. (1924). Zur Axiomatik der speziellen Relativitätstheorie. *Sitz. ber. preuss. Akad. Wiss. Physik.-math. Kl.,* **12**.

Carnap, R. (1928). *Scheinprobleme in der Philosophie.* Engl. transl. in *The Logical Structure of the World.* Berkeley and Los Angeles: University of California Press, 1967.

Carnap, R. (1936). Testability and meaning. *Phil. Sci.* **3**: 419–471; **4**: 1–40.

Carnap, R. (1939). *Foundations of Logic and Mathematics.* Chicago: University of Chicago Press.

Carnap, R. (1942). *Introduction to Semantics.* Cambridge, Mass.: Harvard University Press.

Carnap, R. (1947). *Meaning and Necessity.* Chicago: University of Chicago Press, Enlarged ed. 1956.

Carnap, R. (1950). Empiricism, semantics, and ontology. *Rev. Intern. de Philos.* **4**: 20–40. Repr. in Linsky, Ed. (1952).

Carnap, R. (1952). Meaning postulates. *Phil. Studies* **3**: 65–73.

Carnap, R. (1956). The methodological character of theoretical concepts. In H. Feigl and M. Scriven, Eds., *Minnesota Studies in the Philosophy of Science*, I: 38–76. Minneapolis: University of Minnesota Press.

Carnap, R. (1958). *Introduction to Symbolic Logic and its Applications.* New York: Dover Publications, Inc.

Carnap, R. (1961). On the use of Hilbert's ε-operator in scientific theories. In Bar-Hillel *et al.*, Eds. (1966), 156–164.

Carnap, R. (1963a). Intellectual autobiography. In Schilpp Ed. (1963), pp. 3–84.

Carnap, R. (1963b). Replies and systematic expositions. In Schilpp Ed. (1963), pp. 859–1013.

Castonguay, C. (1972). *Meaning and Existence in Mathematics.* Wien-New York: Springer-Verlag.

Chang, C. C. and H. J. Keisler (1966). *Continuous Model Theory.* Princeton, N.J.: Princeton University Press.

Chang, C. C. and H. J. Keisler (1973). *Model Theory.* Amsterdam-London: North-Holland Publ. Co.

Church, A. (1951). A formulation of the logic of sense and denotation. In P. Henle, H. M. Kallen and S. K. Langer, Eds., *Structure, Method and Meaning: Essays in Honor of Henry M. Scheffer*, pp. 3–24. New York: The liberal Arts Press.

Church, A. (1973/74). Outline of a revised formulation of the logic of sense and denotation. *Nous* **7**: 24–33, **8**: 135–156.

Cole, M. and I. Maltzman, Eds. (1969). *A Handbook of Contemporary Soviet Psychology.* New York and London: Basic Books, Inc.

Davidson, D. (1967). Truth and meaning. *Synthese* **17**: 304–333. Repr. in J. W. Davis, D. J. Hockney, and W. K. Wilson, Eds., *Philosophical Logic.* Dordrecht: D. Reidel Publ. Co., 1969.

DeWitt, B. S. (1970). Quantum mechanics and reality. *Phys. Today* **23**, No. 9, 30–35.

Dingler, H. (1907). *Grundlinien einer Kritik und exakten Theorie der Wissenschaften, insbesondere der Mathematik.* München: Ackermann.

Dummett, M. (1973). *Frege. Philosophy of Language.* London: Duckworth & Co. Ltd.

Dirac, P. A. M. (1942). The physical interpretation of quantum mechanics. *Proc. Roy. Soc. (A)* **180**: 1–40.

Eberle, R. A. (1969). Denotationless terms and predicates expressive of positive qualities. *Theoria* **35**: 104–123.

Einstein, A. (1936). Physics and reality. *J. Franklin Institute* **221**: 313–347.

Enriques, F. (1943). *Problems of Science.* La Salle, Ill.: Open Court.

Everett III, H. (1957). "Relative state" formulation of quantum mechanics. *Rev. Mod. Phys.* **29**: 454–462.

Feigl, H. (1958). The 'mental' and the 'physical'. In H. Feigl, M. Scriven, and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science*, II: 370–497.

Feller, W. (1968). *An Introduction to Probability Theories and its Applications* I, 3rd ed. New York and London: John Wiley & Sons.

Ferrater Mora, J. (1967). *El ser y el sentido.* Madrid: Revista de Occidente.

Feyerabend, P. K. (1970). Against method: Outline of an anarchistic theory of knowledge. In Radner and Winokur Eds. (1970) pp. 17–130.

Fine, A. (1968). Logic, probability, and quantum mechanics. *Phil. Sci.* **35**: 101–111.

Fréchet, M. (1939). The diverse definitions of probability. *Erkenntnis* **8**: 7–23.

Frege, G. (1879). Begriffsschrift. Repr. in *Begriffsschrift und andere Aufsätze.* I. Angelelli, Ed. Hildesheim: Georg Olms 1964.

Frege, G. (1891). Funktion und Begriff. In Angelelli (1967).

Frege, G. (1892). Über Sinn und Bedeutung. In Angelelli (1967).

Frege, G. (1894). Rezension von: E. G. Husserl, Philosophie der Arithmetik. I. In Angelelli (1967).

Frege, G. (1912). Anmerkungen zu P. E. B. Jourdain, The development of the theories of mathematical logic, etc. In Angelelli (1967).

Frege, G. (1969). *Nachgelassene Schriften.* H. Hermes, F. Kambartel and F. Kaulbach, Eds. Hamburg: F. Meiner.

Freudenthal, H. (1970). What about the foundations of physics? *Synthese* **21**: 93–106.

Freudenthal, H. (1971). More about *Foundations of Physics. Foundations of Physics* **1**: 315–323.

Gentilhomme, Y. (1968). Les ensembles flous en linguistique. *Cahiers de linguistique théorique et appliquée* **5**: 47–65.

Giles, R. (1964). *Mathematical Foundations of Thermodynamics.* London: Pergamon.

Goguen, J. A. (1969). The logic of inexact concepts. *Synthese* **19**: 325–373.

Conseth, F. (1938). *La méthode axiomatique.* Paris: Gauthier-Villars.

Goodman, N. (1960). The way the world is. *Rev. Metaphys.* **14**: 48–56.

Goodman, N. (1968). *The Languages of Art.* Indiana and New York: Bobbs Merrill Co.

Goodstein, R. L. (1968). Existence in mathematics. *Compositio mathematica* **20**: 70–82.

Goodstein, R. L. (1969). Empiricism in mathematics. *Dialectica* **23**: 50–57.

Grossberg, S. (1969). Some networks that can learn, remember, and reproduce. *I. J. Math. and Mechanics* **19**: 53–91.

Harrison, M. E. (1965). *Introduction to Switching and Automata Theory.* New York: McGraw-Hill Book Co.

Hartnett, W. E. (1963). *Principles of Modern Mathematics*, 1. Chicago: Scott, Foresman, and Co.

Hartnett, W. E. (1970). *Principles of Modern Mathematics*, 2. Chicago: Scott, Foresman, and Co.

Heisenberg, W. (1955). The development of the interpretation of the quantum theory. In W. Pauli, L. Rosenfeld, and V. Weisskopf, Eds., *Niels Bohr and the Development of Physics*, pp. 12–29. London: Pergamon.

Helmholtz, H. v. (1873). *Lectures on Scientific Subjects.* London: Longmans, Green and Co.

Hempel, C. G. (1970). On the "standard conception" of scientific theories. In M. Radner and S. Winokur, Eds., (1970) pp. 142–163.

Henkin, L., P. Suppes and A. Tarski, Eds. (1959). *The Axiomatic Method.* Amsterdam: North-Holland Publ. Co.

Hermes, H. (1963). *Einführung in die mathematische Logik.* Stuttgart: B. G. Teubner.

Hesse, M. (1965). The explanatory function of metaphor. In Bar-Hillel, ed. (1965), pp. 249–259.

Hilbert, D. (1924). Grundlagen der Physik. In *Gesammelte Abhandlungen*, vol. 3. Berlin:

Julius Springer, 1935.

Hilbert, D. and P. Bernays (1968). *Grundlagen der Mathematik* I, 2nd ed. Berlin-Heidelberg-New York: Springer-Verlag.

Hintikka, J. (1962). *Knowledge and Belief.* Ithaca, N.Y.: Cornell University Press.

Hintikka, J. (1969). *Models for Modalities.* Dordrecht: D. Reidel Publ. Co.

Hooker, C. A. (1974). Systematic realism. *Synthese* **26**: 409–497.

Jost, R. (1965). *The General Theory of Quantized Fields.* Providence, R. I.: Amer. Mathematical Society.

Kalish, D. and R. Montague (1957). Remarks on descriptions and natural deduction. *Arch. f. Math. Logik u. Grundlagenforschung* **3**: 50–73.

Kalmár, L. (1967). Foundations of mathematics – whither now? In I. Lakatos, Ed., *Problems in the Philosophy of Mathematics*, pp. 187–194. Amsterdam: North-Holland Publ. Co.

Katz, J. J. and J. A. Fodor (1963). The structure of a semantic theory. *Language* **39**: 170–210.

Kemeny, J. G. (1956). A new approach to semantics. *J. Symbol. Logic* **21**: 1–27, 149–161.

Kleiner, S. A. (1971). Criteria for meaning changes. In M. Bunge, Ed., *Problems in the Foundations of Physics*, pp. 131–144. New York: Springer-Verlag.

Klemke, E. D., Ed. (1968). *Essays on Frege.* Urbana, Ill.: University of Illinois Press.

Kolmogoroff, A. N. (1963). The theory of probability. In A. D. Aleksandrov, A. N. Kolmogoroff, and M. A. Lavrent'ev, Eds., *Mathematics: Its Content, Methods, and Meaning*, 3 vols. Cambridge, Mass.: M.I.T. Press.

Körner, S. (1964). Deductive unification and idealisation. *Brit. J. Phil. Sci.* **14**: 274–284.

Kripke, S. (1959). A completeness theorem in modal logic. *Journal of Symbolic Logic* **24**: 1–15.

Kuhn, T. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Landau, L. D. and E. M. Lifshitz (1958). *Quantum Mechanics.* London: Pergamon Press.

Leibniz, G. W. v. (1703). *Nouveaux essais sur l'entendement humain.* Paris: Flammarion, s.d.

Leibniz, G. W. v. (1714). The Monadology. In *Philosophical Papers and Letters*, Ed. L. E. Loemker, vol. 2. Chicago: University of Chicago Press 1956.

Lenin, V. I. (1909). *Materialism and Empirio-Criticism.* Moscow: Foreign Language Publ. House 1947.

Lewis, C. I. (1929). *Mind and the World Order.* New York: Scribner.

Lewis, C. I. (1944). The modes of meaning. *Philosophy and Phenomenological Research* **4**: 236–249.

Lewis, C. I. (1951). Notes on the logic of intensions. In P. Henle *et al.*, Eds., *Structure, Method and Meaning.* New York: Liberal Arts Press.

Linsky, L., Ed. (1952). *Semantics and the Philosophy of Language.* Urbana: University of Illinois Press.

Lorenzen, P. (1967). Moralische Argumentationen im Grundlagenstreit der Mathematik. In *Tradition und Kritik: Festschrift für R. Zocher*, pp. 219–227. Stuttgart: F. Fromman.

Luria, A. R. (1969). Speech development and the formation of mental processes. In Cole and Maltzmann, Eds. 1969, pp. 121–162.

Łukasiewicz, J. (1913). Logical foundations of probability theory. In *Selected Works*, Ed. L. Borkowski. Amsterdam: North-Holland Publ. Co. 1970.

MacLane, S. and G. Birkhoff (1967). *Algebra.* New York: Macmillan.

Martin, R. M. (1958). *Truth and Denotation.* Chicago: University of Chicago Press.

Martin, R. M. (1959). *Toward a Systematic Pragmatics.* Amsterdam: North-Holland Publ. Co.

Maxwell, J. C. (1871). Remarks on the mathematical classification of physical quantities. *Proc. London Math. Soc.* **3**: 224–233.

Mendelson, E. (1963). *Introduction to Mathematical Logic.* Princeton, N.J.: D. Van Nostrand Co., Inc.

Meyerson, E. (1908). *Identité et réalité.* Paris: Félix Alcan.

Mill, J. S. (1875). *A System of Logic*, 8th ed. London: Longmans, Green and Co. 1952.

Moisil, Gr. C. (1972). *Essais sur les logiques non chrysippiennes.* Bucarest: Editions de l'Académie de la R. S. de Roumanie.

Montague, R. (1968). Pragmatics. In R. Klibansky, Ed., *Contemporary Philosophy* I: 102–122. Firenze: La Nuova Italia Editrice.

Montague, R. (1970). Pragmatics and intensional logic. *Synthese* **22**: 68–94.

Naess, A. (1956). Synonymity and empirical research. *Methodos* **8**: 3–22.

Nagel, E. (1956). *Logic Without Metaphysics.* Glencoe, Ill.: The Free Press.

Nagel, E., P. Suppes, and A. Tarski, Eds. (1962). *Logic, Methodology and Philosophy of Science.* Stanford: Stanford University Press.

Padoa, A. (1901). Essai d'une théorie algébrique des nombres entiers, précédé d'une introduction à une théorie déductive quelconque. *Bibliothèque du Congrès International de Philosophie* **3**: 309–365.

Parkinson, J. H. R. (1968). *The Theory of Meaning.* Oxford: Oxford University Press.

Peano, G. (1921). Le definizioni in matematica. *Periodico di matematiche* **1**: 175–189.

Popper, K. R. (1959). The propensity interpretation of probability. *Brit. J. Phil. Sci.* **10**: 25–42.

Popper, K. R. (1963a). The demarcation between science and metaphysics. In Schilpp, Ed. (1963), pp. 183–226.

Popper, K. R. (1963b). *Conjectures and Refutations.* New York: Basic Books.

Popper, K. R. (1968). Epistemology without a knowing subject. In B. van Rootselaar and J. F. Staal, Eds. (1968), pp. 333–373.

Popper, K. R. (1970). A realist view of logic, physics, and history. In W. Yourgrau and A. D. Breck, Eds. (1970), pp. 1–37.

Popper, K. R. (1972). *Objective Knowledge.* Oxford: Clarendon Press.

Popper, K. R. (1974). Autobiography. In Schilpp, Ed. (1974).

Przelecki, M. (1969). *The Logic of Empirical Theories.* London: Routledge & Kegan Paul.

Putnam, H. (1962). What theories are not. In Nagel, Suppes and Tarski, Eds. (1962), pp. 240–251.

Quine, W. V. (1952). *From a Logical Point of View.* Cambridge, Mass.: Harvard University Press.

Quine, W. V. (1960). *Word and Object.* Cambridge, Mass.: M.I.T. Press.

Quine, W. V. (1964). Implicit definition sustained. *J. Phil.* **60**: 71–74.

Quine, W. V. (1969). *Ontological Relativity and Other Essays.* New York: Columbia University Press.

Quine, W. V. (1970a). Existence. In W. Yourgrau and A. D. Breck, Eds. (1970), pp. 89–98.

Quine, W. V. (1970b). *Philosophy of Logic.* Englewood Cliffs, N.J.: Prentice-Hall.

Quine, W. V. and N. Goodman (1940). Elimination of extra-logical postulates. *J. Symbol. Logic* **5**: 104–109.

Radner, M. and S. Winokur, Eds. (1970). *Minnesota Studies in the Philosophy of Science* IV. Minneapolis: University of Minnesota Press.

Ramsey, F. P. (1931). *The Foundations of Mathematics.* London: Routledge & Kegan Paul.

Rapoport, A. (1966). *Two-Person Game Theory.* Ann. Arbor: University of Michigan Press.

Reichenbach, H. (1944). *Philosophic Foundations of Quantum Mechanics.* Berkeley and Los

Angeles: University of California Press.

Reichenbach, H. (1949). *The Theory of Probability.* Berkeley and Los Angeles: University of California Press.

Rescher, N. (1973). *The Coherence Theory of Truth.* Oxford: Clarendon Press.

Robinson, A. (1963). *Introduction to Model Theory and to the Metamathematics of Algebra.* North-Holland Publ. Co.

Robinson, A. (1965). Formalism 64. In Bar-Hillel, Ed. (1965), pp. 228–246.

Rootselaar, B. van, and J. F. Staal, Eds. (1968). *Logic, Methodology and Philosophy of Science* III. Amsterdam: North-Holland Publ. Co.

Rosen, R. (1970). *Dynamical System Theory in Biology.* New York: Wiley-Interscience.

Rosenfeld, L. (1961). Foundations of quantum theory and complementarity. *Nature* **190**: 384–388.

Rozeboom, W. (1962). The Factual content of theoretical concepts. In H. Feigl and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* III: 273–357. Minneapolis, Min.: University of Minnesota Press.

Russell, B. (1905). On denoting. *Mind N.S.* **14**: 479–493.

Russell, B. (1919a). *Introduction to Mathematical Philosophy.* London: George Allen & Unwin.

Russell, B. (1919b). On propositions: what they are and how they mean. In R. C. Marsh, Ed., *Logic and Knowledge.* London: George Allen & Unwin 1956.

Salt, D. (1971). Physical axiomatics: Freudenthal vs. Bunge. *Foundations of Phys.* **1**: 307–313.

Scheibe, E. (1973). The approximate explanation and the development of physics. In P. Suppes, L. Henkin, A. Joja, Gr. C. Moisil, Eds., *Logic, Methodology and Philosophy of Science IV*, pp. 931–942. Amsterdam: North-Holland.

Schilpp, P. A., Ed. (1963). The Philosophy of Rudolf Carnap. La Salle, Ill.: Open Court: London: Cambridge University Press.

Schilpp, P. A., Ed. (1974). *The Philosophy of Karl R. Popper.* La Salle, Ill.: Open Court.

Schlick, M. (1932/33). Positivism and realism. In Ayer, Ed. 1959.

Scholz, H. (1941). *Metaphysik als strenge Wissenschaft.* Repr.: Darmstadt: Wissenschaftliche Buchgesellschaft 1965.

Scholz, H. (1969). *Mathesis Universalis*, 2nd ed. Darmstadt: Wissenschaftliche Buchgesellschaft.

Scott, D. (1967). Existence and description in formal logic. In R. Schoenman, Ed., *Philosopher of the Century: Essays in Honour of Bertrand Russell*, pp. 181–200. London: Allen & Unwin.

Scott, D. and C. Strachey (1971). *Towards a mathematical semantics for computer languages.* Oxford: Oxford University Computer Lab.

Shoenfield, J. R. (1967). *Mathematical Logic.* Reading, Mass.: Addison-Wesley.

Stenlund, S. (1973). *The Logic of Description and Existence.* Uppsala: Philosophical Studies.

Strawson, P. F. (1950). On referring. *Mind, N.S.* **59**: 320–344.

Strawson, P. F. (1964). Identifying reference and truth-values. *Theoria* **30**: 96–118.

Strawson, P. F. (1971). *Logico-Linguistic Papers.* London: Methuen & Co. Ltd.

Suppes, P. (1957). *Introduction to Logic.* Princeton, N.J.: D. Van Nostrand.

Suppes, P. (1959). Measurement, empirical meaningfulness, and three-valued logic. In C. West Churchman and P. Ratoosh, Eds., *Measurement: Definitions and Theories.* New York: Wiley.

Suppes, P. (1961). A comparison of the meaning and uses of models in mathematics and the empirical sciences. In H. Freudenthal, Ed., *The Concept and the Role of the Model in Mathematics and Natural and Social Sciences.* Dordrecht: D. Reidel Publ. Co.

Suppes, P. (1965). Logics appropriate to empirical theories. In Addison *et al.*, Eds. (1965), pp. 364–375.

Suppes, P. (1967). *Set-theoretical Structures in Science.* Stanford University: Institute for Mathematical Studies in the Social Sciences.

Suppes, P. (1969). *Studies in the Methodology and Foundations of Science.* Dordrecht: D. Reidel.

Svenonius, L. (1973). Translation and reduction. In Bunge, Ed. (1973a).

Tarski, A. (1934). Some methodological investigations on the definability of concepts. In Tarski (1956).

Tarski, A. (1936). Der Wahrheitsbegriff in den formalisierten Sprachen. In Tarski 1956.

Tarski, A. (1954). Contributions to the theory of models I, II. *Indagationes Mathematicae* **16**: 572–588.

Tarski, A. (1955). Contributions to the theory of models III. *Indagationes Mathematicae* **17**: 56–64.

Tarski, A. (1956). *Logic, Semantics, Metamathematics.* Oxford: Clarendon Press.

Tarski, A., A. Mostowski, and R. M. Robinson (1953). *Undecidable Theories.* Amsterdam: North-Holland Publ. Co.

Truesdell, C. and R. Toupin (1960). The classical field theories. In S. Flügge, Ed., *Handbuch der Physik* III/1. Berlin-Göttingen-Heidelberg: Springer-Verlag.

Tuomela, R. (1973). *Theoretical Concepts.* Wien and New York: Springer-Verlag.

Waissman, F. (1955). Verifiability. In A. Flew, Ed., *Logic and Language.* Oxford: Basil Blackwell.

Wang, H. (1951). Arithmetic translation of axiom systems. *Trans. Amer. Math. Soc.* **71**: 283–293.

Wang, H. (1966). Process and existence in mathematics. In Bar-Hillel *et al.*, Eds. (1966), pp. 328–351.

Weyl, H. (1919). *Raum-Zeit-Materie*, 3rd ed. Berlin: J. Springer.

Wheeler, J. A. (1957). Assessment of Everett's "relative state" formulation of quantum theory. *Rev. Mod. Phys.* **29**: 463–465.

White, H. J. and S. Tauber (1970). *Systems Analysis.* Philadelphia, Pa.: W. B. Saunders.

Whitehead, A. N. (1898). *A Treatise on Universal Algebra* I. Cambridge: Cambridge University Press.

Whitehead, A. N. and B. Russell (1927). *Principia Mathematica* I, 2nd ed. Cambridge: Cambridge University Press.

Williams, D. C. (1937). The realistic interpretation of scientific sentences. *Erkenntnis* **7**: 169–178, 375–382.

Williams, D. C. (1966). *Principles of Empirical Realism.* Springfield, Ill.: Charles C. Thomas.

Winograd, T. (1972). *Understanding Natural Language.* New York: Academic Press.

Yourgrau, W. and A. D. Breck, Eds. (1970). *Physics, Logic, and History.* New York and London: Plenum Press.

Zinov'ev, A. A. (1973). *Foundations of the Logical Theory of Scientific Knowledge* (Complex Logic). Dordrecht: D. Reidel Publ. Co.

INDEX OF NAMES

Ajdukiewicz, Kazimierz 84, 163
Alston, William P. 74
Aristotle 118, 130
Attfield, R. 59
Austin, John Langshaw 162

Bar-Hillel, Yehoshua 59, 134
Barcan-Marcus, Ruth 107
Bell, J. L. 90
Bernays, Paul 12, 153
Beth, Evert W. 11
Birkhoff, Garrett 108, 139
Black, Max 27
Bohr, Niels 182
Bolzano, Bernard 43, 75, 84, 107, 128
Boole, George 8
Bourbaki, Nicholas 136, 168
Brentano, Franz 43
Bunge, Mario 5, 17, 19, 22, 24, 25, 27, 28, 31, 34, 36, 47, 48, 76, 78, 92, 103, 116, 118, 121, 129, 174, 177, 184
Buridan, Jean 43, 84, 195
Burton, W. K. 69

Campbell, Donald T. 179
Carathéodory, Constantin 69
Carnap, Rudolf 7, 11, 30, 44, 69, 70, 72–74, 84, 93, 102, 179, 193, 195
Castonguay, Charles 152, 167
Chang, C. C. 90, 167
Church, Alonzo 84

Davidson, Donald 30
DeWitt, B. S. 15
Dingler, Hugo 43
Dirac, Paul A. M. 38
Dummett, Michael 44
Durrant, M. 59

Eberle, Rolf A. 154
Einstein, Albert 14
Enriques, Federigo 173, 182
Everett III, H. 15, 28

Feigl, Herbert 69
Feller, William 122
Ferrater-Mora, José 194
Feyerabend, Paul K. 63
Fine, Arthur 109
Fodor, J. A. 59
Fréchet, Maurice 34
Frege, Gottlob 31, 43, 44, 52, 75, 83–85, 123, 147, 161, 163, 183
Freudenthal, Hans 14

Gentilhomme, Yves 152
Giles, Robin 69
Goguen, Joseph 151
Gonseth, Ferdinand 138
Goodman, Nelson 173, 183
Goodstein, R. L. 173, 176
Grossberg, Stephen 27

Hanson, Norwood Russell 63
Harrison, Michael E. 16
Hartmann, Nicolai 180
Hegel, Georg Wilhelm Friedrich 118
Heisenberg, Werner 38, 182, 183
Helmholtz, Hermann Ludwig von 179, 182
Hempel, Carl G. 71
Hesse, Mary 27
Hermes, Hans 102
Hilbert, David 12, 78, 84, 153
Hintikka, Jaakko 92, 127, 154
Hooker, Clifford A. 183
Husserl, Edmund 85

Jost, R. 14

Kalish, Donald 154
Kalmár, László 173
Katz, J. J. 59

Keisler, H. J. 90, 167
Kemeny, John G. 8, 13, 38, 84
Kleiner, Scott A. 63
Kolmogoroff, Aleksander N. 3, 122
Körner, Stephan 149, 152
Kuhn, Thomas 63

Landau, Lev 179
Leibniz, Gottfried Wilhelm 43, 85, 92, 140, 172
Lenin, Vladimir Ilich 182
Lewis, Clarence Irving 44, 47, 183
Lifshitz, E. M. 179
Lorenzen, P. 167
Lotze, Rudolf Hermann 75
Łukasiewicz, Jan 105
Luria, A. R. 45

Mac Lane, Saunders 136
Mach, Ernst 173, 179, 193
Martin, Richard 187
Maxwell, James Clerk 66
Meinong, Alexius 117
Mendelson, E. 31
Meyerson, Emile 185
Mill, John Stuart 44
Moisil, Grigore C. 105, 152
Montague, Richard 154, 187
Morris, Charles 186

Naess, Arne 74
Nagel, Ernest 139

Ockham, William of 43

Padoa, Alessandro 174, 176
Parmenides 107
Peano, Giuseppe 174
Peirce, Charles Sanders 43, 186
Piaget, Jean 179, 190
Plato 43, 83–85
Popper, Karl R. 34, 93, 105, 106, 146, 170, 186
Port-Royal school 43
Pota, Kanenas T. 195
Przelecki, Marian 11
Putnam, Hilary 71

Quine, Willard Van Orman 59, 84, 93, 102, 165, 172, 191–192

Radner, M. 63
Ramsey, Frank Plumpton 82
Rapoport, Anatol 34
Reichenbach, Hans 105, 119
Rescher, Nicholas 96
Robinson, Abraham 90, 167
Rosen, Robert 25
Rosenfeld, Leo 39
Rozeboom, William W. 73
Russell, Bertrand 43, 60, 83, 84, 153, 164, 195

Salt, David 177
Scheibe, Erhard 132
Schlick, Moritz 73
Scholz, Heinrich 92, 138, 191
Scott, Dana 121
Shoenfield, Joseph R. 31
Slomson, A. B. 90
Socrates 42, 195
Spencer, Herbert 185
Strachey, C. 121
Strawson, Peter F. 74, 163
Suppes, Patrick 11, 13, 31, 34, 78, 177
Svenonius, Lars 61

Tarski, Alfred 9, 74, 84, 90, 93, 102–103, 174, 179, 195
Tauber, S. 24
Toulmin, Stephen 63
Toupin, Richard 38
Truesdell, Clifford 38

Vienna Circle 31, 43, 73, 193

Waissman, Friedrich 64
Weyl, Hermann 77
Wheeler, John Archibald 28
White, H. J. 24
Whitehead, Alfred North 8
Williams, Donald 44
Winograd, T. 76
Winokur, S. 63
Wittgenstein, Ludwig 31, 43, 72–74, 168, 173

Zinov'ev, A. A. 129

INDEX OF SUBJECTS

A priori 180–181
Abstraction, degree 7
Analogy 27–28, 182
Analyticity 119, 170–173
Antonymy 58
Applicability condition 104
Axiom, semantic 16–18. *See also* Vol. 1, pp. 104–114
Axiomatics 12

Background 177. *See also* Presupposition
Belief 127–128
Biology 39, 179

Coextensiveness 141–142
Coherence doctrine of truth 87–88, 91, 93, 96. *See also* Truth, formal.
Computer 76
Confirmation 97–99, 117–118, 126
Connotation 44. *See also* Sense and Vol. 1, Ch. 5
Consequence 91
Construct XI, 45
  factual 18, 32
  interpretation 1
  theoretical 56
Convention 90
Conventionalism 68
Coordinate transformation 77–80
Correspondence doctrine of truth 88–89
Covariance 76–80

Data 96
Definiteness of significance 55
Definition 174–177
Denotation 44. *See also* Reference and Vol. 1, Ch. 2
Description, definite 153–165
  elementary analysis 155–158
  improper 159
  mathematical analysis 158–161
  meaning of 161–163
  proper 158
Designation 2. *See also* Vol. 1, Ch. 1
  regular 59–60
Domain of validity 146

Elucidation 32–33
Empiricism 69–72
Epistemology 179–189
Equivalence, alethic 116
  extensional 140
  logical 58
  semantical 117
Error, experimental 31, 101
Evidence, empirical 31
Exactification 32–34
Existence 153–155, 191–194
  conceptual 192–193
  neutral 193
  physical 192–193
Explanation, partial 132
Extension 133–146
  algebra of 142–144
  function 136
  lax 150–151
  strict 135
Extensionalism 47, 167–169

Fact 124
Factual domain 18–21
Fictionalism 75
Formalism, mathematical 13, 19–20, 26, 39
  semantical 14, 68

Idealism 82, 84
Inclusion, extensional 141
Indefiniteness, structural 152–153
Inference, scientific 117–118
Information theory 28. *See also* Vol. 1,

pp. 135–138
Intension 144–146. *See also* Vol. 1, Ch. 4
Interpretation 1–41
  adventitious 27, 36
  degree 7, 23
  epistemic 1
  extensional 8
  factual 18–34
  function 5, 18–21
  intensional 8
  literal 27
  maps 18–21
  mathematical 3–13
  metaphorical 27
  nonstandard 3
  partial 7, 24–26
  pragmatic 35–37
  process 37–40
  semantic 35
  semiotic 1–2
  standard 3
Inter-statement confrontation 94–96, 100–101
Invariance 77
Inverse law 144–146
Isomorphism 93–94

Knowledge 180–186

Language, interpreted 2. *See also* Vol. 1, pp. 8–11
  metaphysical neutrality 189–190
  uninterpreted 3. *See also* Theory, abstract
Lattice 4–5
Liar paradox 164
Lindenbaum algebra of predicates 144
Logic 2, 81, 107–108, 118, 134, 170–179, 181
  doxastic 127–128
  epistemic 127–128
  many-valued 106
  metaphysical neutrality 190–194
  modal: look elsewhere

Magnitude 101
Materialism 183
  naive 84–85
Mathematics 3–13, 166–170, 172
Meaning XII, 30, 40–41, 42–80, 85–86, 95–96, 125–126
  addition 48–49
  analysis 42
  assumption 68–72. *See also* Vol. 1, pp. 104–114
  change 62–64
  complement 48–49
  core 148–149
  empirical 64–72
  factual 64–72
  function 47
  inexact 149
  invariance 59–62
  null 49
  objective 75–80
  postulate 71–72
  pragmatic 74–75
  predicate 65
  product 48–49
  space 48
  views on 43–44
Metaphysics 17, 139, 189–195
Model 5–7, 91–92, 147
  extensional 8–13
  factual 21
  full 7
  intensional 8–13
  mathematical 19
  partial 7, 21–24
  set 92
  theory 9–13, 88, 90–93, 147

Nominalism 136. *See also* Vol. 1, pp. 20–22 and 29–32

Objectivity 79, 184
  methodological 169–170
  semantical 169–170
Open texture 64
Operationism 28, 69

Partial order theory 6–7
Platonism 82–85, 136
Possibility 130
Possible world 92, 138–139
Pragmatics 186–189

Pragmatism 74–75, 188
Presupposition 177–179
  meaning-wise 177–178
  methodological 178
  proof-theoretic 177–178
Predicate 123, 135. *See also* Vol. 1, pp. 15–17
Primitive 6
Probability 33–34, 106, 122–125
Propensity 32–34
Proposition 59–60, 85–86, 128–129
Propositional attitude 128
Psychology 128, 179

Quantum theory 28, 36
Quasitautology 173

Rashevsky-Turing morphogenesis theory 25
Realism 70, 185–186
  critical 182, 185
  naive 185
Reference XI–XII, 16–18, 45–49, 87, 134
  adequate 139
  vicarious 51, 53–55
Refutation 97, 118
Representation 66–67, 182–183. *See also* Vol. 1, Ch. 3

Satisfaction 90–91
Semantics XI–XII, 13, 194–195
Semimodel 8
Sense XI–XII, 16–18, 45–49, 87, 167. *See also* Vol. 1, Ch. 5
  minimal 4, 6
  vicarious 51, 53–55
Set theory 134, 168
Sign 1–3
Significance 50–56
  analysis 54
  definiteness 55
  function 50
  likeness 59
Statement 93, 106. *See also* Proposition
  species of 171
Subject, knowing 186–189
Substitution interpretation 107
Switching circuits theory 16–17
Syncategorematic sign 52
Synonymy 57–60
  partial 59
  pragmatic 74
Synthetic proposition 171
Synthetic theory of meaning 43–80
System, semantical 30

Tarski's theory of truth 102–103, 176. *See also* Model theory
Tautology 48, 173
Test, empirical 28, 30–31
Testability 73–74
Theory, abstract 2–6
  concrete 4–5. *See also* Specific theory
  factual 24–25
  fundamental 22
  generic 24–25
  of a model 6, 19
  scientific 13–15, 37, 40
  specific 4
Time 129–130
Topologies of the statement space 110–113
Translation 59–62
  global 61
  pointwise 61
Truth 29–32, 77, 81–132, 140
  bearers 82–86
  condition 30–31, 101–104
  degrees of 105–132
  formal 81, 167
  kinds 82–90
  of fact 93–97
  of reason 90–93
  partial 96–97
  pragmatic concept 120
  value 73–74, 86–87, 97–101, 113–118
  value gap 121
  views on 82–86

Understanding 75–76
Uniqueness 153–161
Use of terms 74

Vagueness 147–153
  in extension 150–152
  in meaning 149

Whorf-Sapir hypotheses 189–190

The companion of this book is Volume 1 of the

Treatise on Basic Philosophy

SENSE AND REFERENCE

TABLE OF CONTENTS

PREFACE

INTRODUCTION

1 Designation

2 Reference

3 Interpretation

4 Intension

5 Gist and content

INDEX OF NAMES

INDEX OF SUBJECTS

BIBLIOGRAPHY

Mario Bunge

Treatise on Basic Philosophy

Volume 3

D. REIDEL PUBLISHING COMPANY / DORDRECHT-HOLLAND / BOSTON-U.S.A.

MARIO BUNGE
McGill University, Montreal, Canada

Treatise on Basic Philosophy – Volume 3

Ontology I:

THE FURNITURE OF THE WORLD

The *Treatise on Basic Philosophy* expounds the first philosophical synthesis to have been conceived after the end of what has been called the Age of Analysis. Unlike many a system in traditional philosophy, Bunge's is formulated in a clear fashion and with the assistance of formal tools such as elementary modern algebra. Besides, it is science-oriented and its author hopes it may help understand science as well as freeing scientific research from obsolete philosophical maxims. In sum, the *Treatise* presents a new system of philosophy in tune with our time.

The first two volumes propose a new semantical theory and, indeed, the only existing theory that elucidates and interrelates all of the key semantic concepts met with in the analysis of ordinary knowledge and science. In particular, these volumes present original theories of reference (as distinct from extension), of intensions, of meaning, and of truths of fact. Some use is made of elementary set-theoretic and algebraic concepts, such as those of lattice and Boolean algebra. On the other hand modal logics are not employed and model theory is dealt with only tangentially. This work constitutes the first detailed and systematic treatment of the semantics of empirical knowledge (as distinct from the semantics of either mathematics or natural language).

(continued on backflap)

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND / BOSTON-U.S.A.

TREATISE ON BASIC PHILOSOPHY

Volume 3

ONTOLOGY I: THE FURNITURE OF THE WORLD

FROM THE LIBRARY OF
JOHN McCARTHY
(1927 - 2011)

TREATISE ON BASIC PHILOSOPHY

1

SEMANTICS I *Sense and Reference*

2

SEMANTICS II *Interpretation and Truth*

3

ONTOLOGY I *The Furniture of the World*

4

ONTOLOGY II *A World of Systems*

5

EPISTEMOLOGY I *The Strategy of Knowing*

6

EPISTEMOLOGY II *Philosophy of Science*

7

ETHICS *The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 3

Ontology I:

THE FURNITURE OF THE WORLD

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND/BOSTON-U.S.A.

Library of Congress Cataloging in Publication Data

Bunge, Mario Augusto.
The furniture of the world.

(His Ontology; 1) (His Treatise on basic philosophy; v. 3)
Bibliography: p.
Includes indexes.
1. Substance (Philosophy). 2. Form (Philosophy). 3. Change. 4. Space and time. I. Title.
BD311.B84 vol. 1 [BD331] 111s [110] 77-1756
ISBN 90-277-0780-4
ISBN 90-277-0785-5 pbk.

Published by D. Reidel Publishing Company,
P.O. Box 17, Dordrecht, Holland

Sold and distributed in the U.S.A., Canada, and Mexico
by D. Reidel Publishing Company, Inc.
Lincoln Building, 160 Old Derby Street, Hingham,
Mass. 02043, U.S.A.

All Rights Reserved
Copyright © 1977 by D. Reidel Publishing Company, Dordrecht, Holland
No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any informational storage and
retrieval system, without written permission from the copyright owner

Printed in The Netherlands

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and of right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient – except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can

become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: "Do your own thing. Your reward will be doing it, your punishment having done it".

CONTENTS OF *ONTOLOGY I*

GENERAL PREFACE TO THE *TREATISE* V

PREFACE TO *ONTOLOGY I* XIII

ACKNOWLEDGEMENTS XV

SPECIAL SYMBOLS XVI

INTRODUCTION 1
1. Ontological Problems 1
2. The Business of Ontology 3
3. Is Ontology Possible? 6
4. The Method of Scientific Ontology 7
5. The Goals of Scientific Ontology 9
6. Ontology and Formal Science 13
7. The Ontology of Science 15
8. Ontological Inputs and Outputs of Science and Technology 19
9. Uses of Ontology 22
10. Concluding Remarks 24

1. SUBSTANCE 26
   1. Association 26
      1.1. Concatenation and its Ontological Interpretation 26
      1.2. Axiomatic Foundation of Association Theory 28
      1.3. Consequences 31
      1.4. Atom Aggregates 36
      1.5. Clustering 37
      1.6. Historical Remark 38

   2. Assembly 39
      2.1. Juxtaposition and Superposition: Intuitive Idea 39
      2.2. Formalization 40
      2.3. Definitions 42
      2.4. Some Consequences 45

2.5. Atoms and Levels 47
2.6. Alternative Formalizations 49
2.7. Concluding Remarks 50

3. Entities and Sets 50
3.1. The Null Individual and the World 50
3.2. Entities and Concepts 52
3.3. Existence and Individuation 53

4. Concluding Remarks 55

2. FORM 57
1. Property and Attribute 58
1.1. Difference between Property and Attribute 58
1.2. Attribute-Property Correspondence 59

2. Analysis 62
2.1. Property in General and Property of a Particular 62
2.2. Intrinsic and Mutual, Primary and Secondary 65

3. Theory 69
3.1. Unarization and Dichotomization 69
3.2. Basic Assumptions and Conventions 72
3.3. Laws as Properties 77
3.4. Precedence and Conjunction of Properties 80
3.5. Similarity 85
3.6. Indiscernibility 90

4. Properties of Properties 92
4.1. Identity and Difference of Properties 92
4.2. Property Weight 94
4.3. Resultants and Emergents 97
4.4. Properties of Properties 98

5. Status of Properties 99
5.1. The Reality of Properties 99
5.2. A Critique of Platonism 102
5.3. The Problem of Universals 104

6. Concluding Remarks 108

3. THING 110
1. Thing and Model Thing 110
1.1. Thing: Definition 110

1.2. Assumptions 112
1.3. Thing and Construct 116
1.4. Model Thing 119

2. State 123
2.1. Centrality of the State Concept 123
2.2. State Function 125
2.3. Law Statements as Restrictions on State Functions 128
2.4. State Space: Preliminaries 131
2.5. Definition of a State Space 133
2.6. Equivalent Representations of States 136
2.7. State and State Preparation 138
2.8. Concluding Remarks 139

3. From Class to Natural Kind 140
3.1. Classes of Things 140
3.2. Ideals and Filters of Classes of Things 140
3.3. Kinds and Species 143
3.4. The Algebra of Kinds 147
3.5. Variety 150

4. The World 152
4.1. What Does the World Consist in and of? 152
4.2. Individuals, Populations, Communities, and Species 153
4.3. Existence Concepts 155
4.4. Nothingness and Virtual Existence 158
4.5. Existence Criteria 160

5. Concluding Remarks 162

4. POSSIBILITY 164
1. Conceptual Possibility 165
1.1. Possibility Concepts 165
1.2. Four Concepts of Conceptual Possibility 165
1.3. Conceptual Possibility: Relative 168

2. Real Possibility 168
2.1. Fact 168
2.2. Chrysippian Possibility 172
2.3. Real Possibility as Lawfulness 173
2.4. Factual Necessity 174
2.5. Possibility Criteria 177

3. Disposition 179
3.1. Intuitive Idea 179

3.2. Elucidation 181
3.3. Potency and Act 183
3.4. Unrealized Possibilities and Counterfactuals 184

4. Probability 185
4.1. Abstract Concept 185
4.2. Probability State Space 187
4.3. Propensity Interpretation 190

5. Chance Propensity 194
5.1. Irreducible Potentialities 194
5.2. Analysis 197
5.3. Upshot 198

6. Marginalia 199
6.1. Modal Logic and Real Possibility 199
6.2. Possible Worlds Metaphysics 202
6.3. Modality and Probability 204
6.4. Randomness 208
6.5. Probability and Causality 210
6.6. The Many Worlds Interpretation of Quantum Mechanics 211

7. Concluding Remarks 212

5. CHANGE 215
1. Changeability 216
1.1. Preliminaries 216
1.2. Changeability 218

2. Event 221
2.1. The Ordered Pair Representation of Events 221
2.2. The Event Space 222
2.3. The Representation of Processes 226
2.4. The Space of Lawful Events 229
2.5. Keeping Track of Changing States 232
2.6. Rate, Extent, and Change Potential 237

3. Process 243
3.1. Serial Change: Types 243
3.2. General Concepts and Principles 251

4. Action and Reaction 256
4.1. Induced Change 256
4.2. Aggregates and Systems 262
4.3. Reference Frame 264

5. *Panta Rhei* 267
  5.1. Fact 267
  5.2. Dynamicism 268
  5.3. Interconnectedness 270
  5.4. Three Misconceptions 271

6. Concluding Remarks 273

6. SPACETIME 276
1. Conflicting Views 278
  1.1. The Three Main Views 278
  1.2. Approaches to Chronotopics Building 281

2. Space 283
  2.1. Interposition 283
  2.2. A Philosopher's Space 285
  2.3. The Physicist's Space 287
  2.4. Bulk and Shape 293
  2.5. Concluding Remarks 294

3. Duration 296
  3.1. Intuitive Idea 296
  3.2. Before and After 297
  3.3. Duration 301

4. Spacetime 305
  4.1. Spacetime, the Basic Network of Events 305
  4.2. Position in Spacetime 309
  4.3. Change in Spacetime 312

5. Spatiotemporal Properties 314
  5.1. Does Spacetime have any Properties? 314
  5.2. Time Reversal and Process Reversibility 317
  5.3. Antecedence ("Causality") Principle 320
  5.4. Action by Contact 323
  5.5. Spatiotemporal Contiguity 325
  5.6. The Causal Relation 326

6. Matters of Existence 327
  6.1. Existence in Space and Time 327
  6.2. Existence of Space and Time 328

7. Concluding Remarks 330

BIBLIOGRAPHY 334

INDEX OF NAMES 344

INDEX OF SUBJECTS 348

PREFACE TO *ONTOLOGY I*

This book and its companion, namely Volume 4 of our *Treatise*, concern the basic traits and patterns of the real world. Their joint title could well be *The Structure of Reality*. They constitute then a work in ontology, metaphysics, philosophical cosmology, or general theory of systems. Our work is in line with an old and noble if maligned tradition: that of the pre-Socratic philosophers, Aristotle, Thomas Aquinas, Descartes, Spinoza, Leibniz, Hobbes, Helvetius, d'Holbach, Lotze, Engels, Peirce, Russell, and Whitehead. But at the same time it departs from tradition in the matter of method. In fact our aim is to take the rich legacy of ontological problems and hints bequeathed us by traditional metaphysics, add to it the ontological presuppositions of contemporary scientific research, top it with new hypotheses compatible with the science of the day, and elaborate the whole with the help of some mathematical tools.

The end result of our research is, like that of many a metaphysical venture in the past, a conceptual system. It is hoped that this system will not be ridiculously at variance with reason and experience. It is intended moreover to be both exact and scientific: exact in the sense that the theories composing it have a definite mathematical structure, and scientific in that these theories be consistent with and moreover rather close to science – or rather the bulk of science. Furthermore, to the extent that we succeed in our attempt, science and ontology will emerge not as disjoint but as overlapping. The sciences are regional ontologies and ontology is general science. After all, every substantive scientific problem is a subproblem of the problem of ontology, to wit, *What is the world like*?

After a long period underground, talk about metaphysics has again become respectable. However, we shall not be talking at length about ontology except in the Introduction. We shall instead do ontology. In the process we shall attempt to exhibit the mathematical structure of our concepts and we shall make the most of science. Being systematic our ontology may disappoint the historian. Being largely mathematical in form it will be pushed aside by the lover of grand verbal (but sometimes deep and fascinating) systems – not to speak of the lover of petty verbal matters. And being science-oriented it will fail to appeal to the friend of the esoteric. Indeed we shall be concerned with concrete objects such as atoms, fields, organisms, and societies. We shall abstain from talking about items that are neither concrete things nor properties, states or changes thereof. Any fictions entering our system will be devices useful in accounting for the structure of reality. (Constructs were dealt with in Volumes 1 and 2 of this work.)

The first ideas for this work dawned upon me when I was engaged in axiomatizing some basic physical theories involving ontological concepts such as those of thing, property, possibility, change, space, and time, none of which are the exclusive property of physics but all of which belong to the metaphysical background of this science, or protophysics (Bunge 1967b). And the earliest plan for this work occurred to me a bright day of June 1966 when travelling from Freiburg im Breisgau to Geneva at the invitation of Jean Piaget. I have been working on this project ever since, on and off, stimulated by what seemed a grand design and occasionally inhibited by the difficulties met with in carrying it out. The result is a system but not a closed and final one: there is much room for improvement and of course also for divergent developments.

This volume deals with the concepts of substance, form (or property), thing (or concrete object), possibility, change, space, and time. The companion volume, *A World of Systems*, will tackle the concepts of system, novelty, biosystem, psychosystem, and sociosystem.

ACKNOWLEDGEMENTS

It is a pleasure to thank the various generations of curious, bewildered and merciless students who survived my ontology courses at McGill University and the Universidad Nacional Autónoma de México from 1969 to 1976. I am also indebted to Professors Rutherford Aris (Chemical Engineering, Minnesota), Thomas A. Brody (Physics, UNAM, México), Máximo García Sucre (Physics, IVIC, Caracas), Tomás Garza (IIMAS, UNAM, México), Andrés J. Kálnay (Physics, IVIC, Caracas), and Roberto Torretti (Philosophy, Puerto Rico) for critical remarks and suggestions. My former research assistants Dr David Salt and Dr Gerhard Vollmer made a number of useful remarks. My research assistant Mr Robert Blohm enhanced the clarity, improved the grammar, and asked questions still to be answered. But my greatest debt is to my former research associates Professors Adalberto García Máynez (Mathematics, IPN, México), William E. Hartnett (Mathematics, SUNY, Plattsburgh), and Arturo A. L. Sangalli (Mathematics, Ottawa), none of whom understood friendship to be in the way of rigor.

The Alexander von Humboldt-Stiftung covered my first incursions into the intersection of physics and metaphysics (1965-66). The Canada Council contributed to this project by awarding me research grants (1969–72, 1974–76), one of them on behalf of the Killam Foundation. And the John Simon Guggenheim Memorial Foundation awarded me a fellowship that gave me a happy and fruitful year (1972–73). It was during the tenure of that fellowship, at the ETH Zürich, that I started to write this book. I am grateful to all those organizations for their support, and to Professors Gerhard Huber and Peter Huber for their hospitality at the ETH. Last, but not least, I thank my guide in the fascinating and puzzling Mexican labyrinth, Professor Fernando Salmerón, director of the Instituto de Investigaciones Filosóficas of the UNAM, where this volume acquired its final shape during the 1975–76 academic year.

MARIO BUNGE

SPECIAL SYMBOLS

$x \rhd y$	x *acts* on y
$c = a \circ b$	c is the *association* of individuals a and b
$A \times B$	the *cartesian product* of sets A and B
$\langle s, s', g \rangle$	the *change* from state s to state s' along curve g
C	set of constructs
$\mathscr{C}(x)$	the *composition* of thing x
$x \lfloor y$	x and y are *detached*
$E_{\mathbb{L}}(x)$	the *event space* of thing x
$E_A x$	x *exists* in A
$\mathscr{E}(A)$	the *extension* of attribute (predicate) A
$f: A \to B$	*function* f maps set A into set B
F	set of *facts*
$\mathbb{F} = \langle F_1, F_2, \ldots, F_n \rangle$	*state function*
$h(x)$	the *history* of thing x
$a \vert c \vert b$	c *interposes* between a and b
$c = a \dotplus b$	c is the *juxtaposition* of a and b
$k(\mathbb{R})$	the *kind* of things sharing all properties in $\mathbb{R}$
$G_{\mathbb{L}}(x)$	the set of *lawful transformations* of the states of x
$\mathbb{L}(x)$	the *laws* of thing x
$\langle x, y \rangle$	the *ordered pair* of x and y
$\square$	the *null thing*
$x \sqsubset y$	x is a *part* of y
$\mathscr{P}(S) = 2^S$	the *power set* of S
$P \preccurlyeq Q$	property P *precedes* property Q
$\mathbb{P}$	the set of *all properties*
$p(x)$	the collection of *properties of thing* x
Pr	*probability* function
$\mathbb{R}$	the *real line*
$S_{\mathbb{L}}(x)$	*lawful state space* of thing x
$c = a \dot{\times} b$	c is the *superposition* of things a and b
S	the set of *all substantial (concrete) individuals*
$\mathscr{S}(P)$	the *scope of property* P
$[T] = \inf T$	the *additive aggregation* of all things in T
$(T) = \sup T$	the *multiplicative aggregation* of all things in T
Θ	the *set of all things*
▯	the *world* or *universe*

INTRODUCTION

In this Introduction we shall sketch the business of ontology, or metaphysics, and shall locate it on the map of learning. This has to be done because there are many ways of construing the word 'ontology' and because of the bad reputation metaphysics has suffered until recently – a well deserved one in most cases.

1. Ontological problems

Ontological (or metaphysical) views are answers to ontological questions. And ontological (or metaphysical) questions are questions with an extremely wide scope, such as 'Is the world material or ideal – or perhaps neutral?', 'Is there radical novelty, and if so how does it come about?', 'Is there objective chance or just an appearance of such due to human ignorance?', 'How is the mental related to the physical?', 'Is a community anything but the set of its members?', and 'Are there laws of history?'.

Just as religion was born from helplessness, ideology from conflict, and technology from the need to master the environment, so metaphysics – just like theoretical science – was probably begotten by the awe and bewilderment at the boundless variety and apparent chaos of the phenomenal world, i.e. the sum total of human experience. Like the scientist, the metaphysician looked and looks for unity in diversity, for pattern in disorder, for structure in the amorphous heap of phenomena – and in some cases even for some sense, direction or finality in reality as a whole. Metaphysics and science have then the same origin. However, they can be distinguished up to a certain point, namely by the scope of their problems. Whereas the scientific specialist deals with rather specific questions of fact, the ontologist is concerned with all of the factual domains: he is a generalist not a fragmentarian. His enterprise is more ambitious, hence also riskier, than any one scientific project. But the two concerns are not mutually exclusive and, in fact, sometimes they are indistinguishable: an extremely general scientific question may be a special ontological one.

Ontological questions are not easy to characterize or even to recognize as meaningful in isolation from ontological frameworks or ontological theories. Consider for example the following questions:

(i) Why is there something rather than nothing?
(ii) Does essence precede existence?
(iii) What is being?
(iv) Where is one?
(v) What is there?

The first question makes sense in any creationist system of theodicy, such as Leibniz', but it makes no sense elsewhere. The second, at first blush unintelligible, makes good sense in a Platonic metaphysics, where essences are ideal and prior to physical existents. The third question becomes meaningful if reformulated as 'What are the features common to all existents?'. The fourth question was asked by my son Eric when he was 18 months old. It must have made some sense in his own *Weltanschauung* and, rewriting 'one' with a capital O, it might be safely attributed to Parmenides. The fifth – which is how Quine seems to understand the task of ontology – calls for either an exhaustive inventory of existents – a job for scientists of the Baconian persuasion – or a simplistic answer such as 'There are bodies and persons' (Strawson, 1959).

The preceding questions make hardly any sense in the system to be developed in this book. On the other hand the following problems do make sense in it and moreover can be given definite answers:

(vi) Are things bundles of properties? (No.)
(vii) Are there natural kinds? (Yes.)
(viii) Is change possible without an unchanging substrate? (Yes.)
(ix) How do emergent properties come about? (Wait for Volume 4.)
(x) What is the mind? (Ditto.)

These questions are both fundamental and extremely general. Moreover they are factual questions – only, comprehensive or cross-disciplinary rather than special. There are as many such questions as we care to ask. The more we come to know, the more problems we can pose and the less final our solutions prove to be. In both regards then – factual content and open-endedness – ontological questions are no different from scientific ones. (For the latter see Bunge, 1967a, Vol. I, Ch. 4.) They differ only in scope. And even this difference is often non-existent, as will be seen in Sec. 7.

2. THE BUSINESS OF ONTOLOGY

At least the following ten conceptions of the concern of ontology (or metaphysics) have followers nowadays:

(i) Metaphysics is a discourse (in either ancient Greek or modern German) *on Being, Nothingness, and Dasein* [human existence] (Heidegger, 1953). Objection: impossible, because the said discourse is unintelligible and moreover avowedly irrational. If in doubt try to read Heidegger or Sartre.

(ii) Metaphysics is a collection of *instinctive* (as opposed to intellectual) *beliefs* (Bergson, 1903). Objection: if metaphysics is to be a discipline then it cannot collect blindly any received ideas, be they "instinctive" or gotten from tradition. The fact that cave men held "instinctive" beliefs without subjecting them to methodic criticism does not justify us in upholding an attitude that would never have taken us out of the cave.

(iii) Metaphysics is the *justification of instinctive beliefs*: "the finding of reasons, good, bad or indifferent, for what we believe on instinct" Strawson, 1959, p. 247). Objection: our least educated tenets must surely be studied – by the cultural anthropologist. Scholars, whether scientists or metaphysicians, are supposed to examine, refine or reject that starting point – and above all to propose new ideas. Reassuring the cave man in his primitive metaphysics is worse than sharing it.

(iv) Metaphysics is "*the science of absolute presuppositions*" (Collingwood, 1940). That is, metaphysics is the study of all presuppositions, of any discipline, in so far as they are absolute, i.e. lurk behind every question and every answer, and are moreover beyond question. This is a respectable view. However, it is open to the following objections: (*a*) most presuppositions are not absolute but are bound to rise and fall with the special theory concerned: look at the history of ideas; (*b*) even though metaphysics does study some of the presuppositions of science, it does not handle them all, as some of them are purely formal (logical or mathematical) and others are methodological.

(v) Metaphysics deals with *everything thinkable*, whether or not it actually exists, whether reasonable or absurd: it is concerned with "the totality of the objects of knowledge" (Meinong, 1904, in Chisholm, 1960, pp. 78–79). Objections: (*a*) no theory is possible that will encompass both concrete objects and conceptual ones; in particular, logical truths may refer to anything but they do not describe or represent any objects except the logical concepts ("or", "all", etc.); (*b*) objects known to be fantastic, such as Pegasus, may be imagined but cannot constitute the subject of any discipline: only our beliefs about such mythological objects can be studied scientifically.

(vi) Metaphysics is *the study of objects neither physical nor conceptual* – i.e. of spiritual beings, and of God and his celestial court in the first place. This opinion is quite popular and was occasionally voiced by Thomas himself (1259 Bk. I, Ch. IV). Objection: that is the proper subject of theology, which is no longer recognized as a part of philosophy.

(vii) Metaphysics is *the science of being as such*: unlike the special sciences, each of which investigates one class of being, metaphysics investigates "all the species of being *qua* being" and "the attributes which belong to it *qua* being" (Aristotle, *Metaphysics* Bk. IV, Chs. 1 and 2). This is what nowadays one would call *general* ontology by contrast to the various *special* or *regional* ontologies (of the biological, the social, etc.). Certainly the Philosopher had a correct grasp of the relation between metaphysics (general) and the sciences (special). Still, the following objections must be raised: (*a*) the formulation is too imprecise, so much so that it has suggested to some that becoming is not within the purview of metaphysics – an opinion certainly not shared by the Stagirite, who was centrally concerned with change; (*b*) a science of pure being is a contradiction in terms because it has no definite subject matter (Collingwood, 1940, pp. 10–11).

(viii) Metaphysics is *the study of change*: of events and processes – because this is what things are (Whitehead, 1929). Objection: an event is a change in the condition (state) of some thing and therefore cannot be studied apart from it any more than things can be studied apart from their changes.

(ix) Metaphysics concerns *all possible worlds*: it is an ontological interpretation of logic. A system of metaphysics is a set of statements satisfying two conditions: (*a*) "The horizon [set of referents] of a significant metaphysical statement must surpass in an unambiguous way the horizon of a physical statement", and (*b*) "A metaphysical statement must not lag behind a physical statement as far as exactness and stability [*Standfestigkeit*] are concerned" (Scholz, 1941, pp. 138–139). While I have no quarrel with the exactness condition, I dispute the others. My objections are: (*a*) the fact that logic may refer (apply) to anything does not render it a theory of all possible worlds; (*b*) while some metaphysical statements do concern all concrete things, others refer to things belonging to certain genera such as physical objects or organisms or societies; (*c*) metaphysical statements cannot be any less fallible than scientific (e.g. physical) statements.

(x) Metaphysics is *general cosmology or general science*: it is the science concerned with the whole of reality – which is not the same as reality as a whole. "Its business is to study the most general features of reality and real objects" (Peirce, 1892–93, p. 5). It "is concerned with all questions of a general and fundamental character as to the nature of the real" (Montagu, 1925, p. 31; see also Woodger, 1929; Williams, 1937; and Quinton, 1973). In other words, metaphysics studies the generic (nonspecific) traits of every mode of being and becoming, as well as the peculiar features of the major genera of existents. This is the task Hegel (1812–16) assigned to "objective logic" and Engels (1878) to what came to be known as dialectical materialism.

We adopt the latter position: we maintain that the ontologist should stake out the main traits of the real world as known through science, and that he should proceed in a clear and systematic way. He should recognize, analyze and interrelate those concepts enabling him to produce a unified picture of reality. (The word 'reality' is here understood in a strict and non-Platonic sense, namely as the concrete world.) In this sense the reader is real and so is any utterance of the word 'reader'; but the concept designated by this word is unreal.

Because unreal objects have nonphysical properties, they satisfy nonphysical laws if any. For this reason it is impossible to make any nontautological statements applying to all objects: ontology, as conceived by Meinong and Leśniewski, i.e. as a general theory of objects of any kind, and yet different from logic, is impossible. So is the modern version of this doctrine, namely general systems theory construed as a mathematical theory "dealing with the explanations of observed phenomena or conceptual constructs in terms of information-processing and decision-making concepts" (Mesarović in Klir, 1972, p. 253). If the "system" is purely conceptual, as is the case with a number system, then it cannot combine with material systems to form supersystems, it cannot interact with them, it does not obey laws of the same kind and therefore it cannot be studied with the special methods of factual science. Whatever we may think of the construct-thing duality, whether we take a Platonist or a materialist stand, or whether or not we wish to reduce objects of one kind to objects of another, we should keep the duality at the methodological level. (Recall Vol. 1, Ch. 1, Sec. 3.)

We leave formal science, i.e. logic, mathematics, and semantics, the task of studying (and creating) formal or ideal objects of the law-abiding kind, such as sets and categories. (More in Sec. 6.) We take factual (natural or social) science and ontology to be the only disciplines concerned with concrete objects. And we assign ontology the task of constructing the most general theories concerning such and only such objects. But is this task possible?

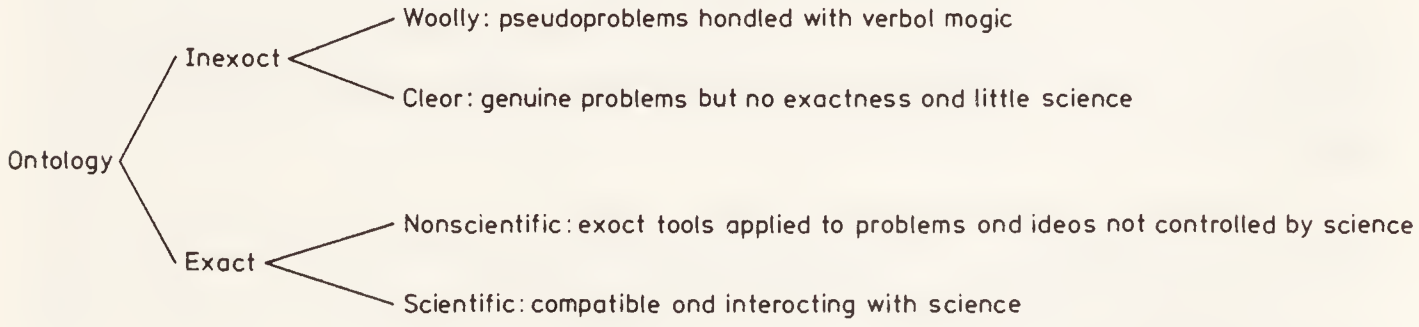
3. IS ONTOLOGY POSSIBLE?

Up until the first scientific revolution, in the early 17th century, metaphysics was taken for granted and was usually regarded as an adjunct to theology. Thereafter the formidable successes of science, jointly with the failure of the grand metaphysical systems – chiefly those of Aristotle, Aquinas, Descartes and Leibniz, and later on Hegel's and Heidegger's – suggested, nay "proved", that metaphysics had become impossible and had been replaced with science.

Those who deny the possibility of metaphysics as an authentic and creative intellectual endeavor hold that metaphysical sentences are either meaningless (hence neither true nor false) or meaningful but untestable; or, if meaningful and testable, then false. Or they hold that, aside from questions of meaning and truth, such statements have no value either practical or heuristic. That many metaphysical statements – e.g. most of Hegel's and all of Heidegger's – are hardly meaningful, is true and sad enough but speaks only against bad metaphysics. (Likewise the fact that most statements about the mental are nonscientific does not disqualify psychology as a science.) That metaphysical propositions cannot be checked for truth the same way scientific statements are tested, is correct but in no way proves that they are beyond criticism and confirmation. That many, perhaps most metaphysical propositions have turned out to be false, and many more will be refuted, is again true but does not render metaphysics any less possible than science, which is in the same predicament. Finally, it is not quite correct that metaphysics lacks in heuristic or even practical import. For one thing scientific research makes use of a number of ontological hypotheses, as will be seen in Sec. 7. For another the philosophic core of every world view and every ideology is, for better or (usually) worse, some ontological system or other together with some value system. In short, it is not true that ontology has become impossible after the birth of modern science. What has become impossible *de jure* – though unfortunately not *de facto* – is nonscientific ontology.

The sceptic may not deny the possibility of ontology but will regard it as inconclusive. With this we concur: metaphysics is indeed inconclusive – but so is factual science. Genuine knowledge is characterized not by certainty but by the possibility of correction in an endless effort at attaining truth and depth. This holds for ontological knowledge no less than for scientific knowledge. Let the sceptic criticize every single metaphysical system: let us keep him busy by offering him new metaphysical theories.

Ontology is not just possible but very much alive these days. (As a matter of fact it has never ceased to exist: it just went underground for a while.) A casual look at the literature will bear out this contention. Moreover, metaphysics only recently has undergone a revolution so deep that nobody has noticed it: indeed ontology has gone mathematical and is being cultivated by engineers and computer scientists. As a matter of fact a number of technologists have developed, over the past three decades, certain exact theories concerning the most basic traits of entities or systems of various genera. Switching theory, network theory, automata theory, linear systems theory, control theory, mathematical machine theory, and information theory are among the youngest metaphysical offspring of contemporary technology (Bunge, 1971, 1974d, 1977a). This kind of ontology, both exact and scientific, is the one we wish to develop and systematize. Its place in the map of metaphysics is shown in the following table. (Nonacademic varieties of metaphysics, mostly esoteric, have been left aside.)


Ontology
Inexact
Woolly: pseudoproblems handled with verbal magic
Clear: genuine problems but no exactness and little science
Exact
Nonscientific: exact tools applied to problems and ideas not controlled by science
Scientific: compatible and interacting with science

4. THE METHOD OF SCIENTIFIC ONTOLOGY

The ontology we wish to build is both exact and contiguous with science. ‘Exact’ means of course logical or mathematical in form. Exact metaphysics is then the set of metaphysical theories built with the explicit help of logic or mathematics. For example, a mathematical theory of the synthesis of wholes out of lower level units would qualify as a piece of exact metaphysics. Not so a theory, however interesting, tackling the same problem in ordinary knowledge terms: it would remain at the level of inexact metaphysics because of the ambiguity and imprecision of ordinary language.

We are interested in the scientific variety of exact metaphysics, that is, in the ontological theories that, in addition to being exact, are scientific. The difference between 'exact' and 'scientific' is this: the latter implies the former but the converse is false. That is, there are systems of exact metaphysics that are out of touch with factual (natural and social) science. For example, Leibniz, Bolzano, Scholz and Montague were exact metaphysicians but they conceived of metaphysics as an a priori science (see e.g. Bolzano, 1849, p. 29); hence their work is not in tune with the science of their day. Likewise most of the contemporary essays on possible worlds, temporal logic, and causality, though often exact, are far removed from science and sometimes even incompatible with it. There is little if any philosophical meat in these theories; and when there is some, it turns out to be stale.

As for scientific metaphysics, it is still largely a program. Even Peirce (1892–93), perhaps the first to employ the expression, did not advance beyond some programmatic remarks. To implement that program we need some guidelines. Here are some of the *regulae philosophandi more geometrico et scientifico* we shall try to abide by:

R1 *Take the metaphysical tradition into account but do not stick to it*: go ahead revising the traditional stock of problems and solutions, posing new problems, and trying new solutions to both old and new questions.

R2 *Avoid any words that fail to convey clear ideas*: obscurity is not the mark of profundity but of confusion or even of intellectual swindle. As for fuzzy ideas – all ideas are fuzzy when newly born – try and refine them.

R3 *Try and formalize everything*: whatever is worth saying in any theoretical discipline, including metaphysics, can be said with the help of mathematics and ought so to be said for the sake of clarity and systematizability.

R4 *Do not mistake symbolization for mathematization*: stenography does not elucidate and does not systematize. Unless a construct is assigned a definite mathematical status (as set, relation, function, group, topological space, or what have you) it is not exact and may be a fake, i.e. a *flatus vocis* rather than a genuine concept.

R5 *Strive for rigor but do not allow it to curtail vigor*: exactness is a means not an end – a means to attain clarity, systemicity, cogency and testability. Insisting on rigor for its own sake and at the price of giving up deep intuitions is a mark of sterility.

R6 *Explain the concrete by the abstract* rather than the other way around. Invert Russell's advice to replace "inferred" (unobserved, hypothetical) entities with logical constructions out of sense impressions. Imitate the atomists and the field theorists. Admit observable, ostensive and secondary properties only if they can be analyzed in terms of primary properties.

R7 *Keep clear of subjectivism* – e.g. avoid definitions in terms of personal experience. Example of the strategy to be avoided: Whitehead (1920, p. 13) defined the concept of a fact as "the undifferentiated terminus of sense-awareness". Every ontological elucidation, even of concepts concerning subjective phenomena, should be couched in purely objective (subject-free) terms.

R8 *Do not reify whatever is not a thing*, and do not treat as an autonomous entity what is but the result of abstraction. E.g. do not talk about events apart from or even as constituting changing things.

R9 *Strive for systemicity*: try to build theories and establish links among them. Do not introduce – except by way of extrasystematic remarks – concepts that have not yet been elucidated: proceed in orderly fashion, and if necessary axiomatically. Philosophical analysis is indispensable but not enough, and is best done in the context of a system anyway.

R10 *Check* your metaphysical hypotheses and theories not only for internal consistency but also for compatibility and even contiguity with contemporary science.

Abiding by these rules should help us in building science-oriented ontological frameworks and even theories, though it won't secure them.

5. The goals of scientific ontology

Every philosophical activity has two possible intellectual goals: analysis or synthesis – i.e. taking apart or building. Philosophical analysis examines certain concepts and propositions in order to clarify them. Philosophical synthesis creates frameworks and theories in order not only to elucidate notions and statements but also to understand what goes on in the world. The two operations are complementary rather than mutually exclusive. In fact the best analysis is that performed in a definite theoretical context, just as the best synthesis is that which articulates analyzed notions. In any case it is synthesis, not analysis, that provides some understanding of a field of reality or of human experience. Thus only a theory of ideation, rather than an analysis of the notion of ideation – let alone an analysis of the word 'ideation' as used in a given community – can account for ideation.

In the case of ontology, analysis bears on any metaphysical concepts or propositions – or candidates for either role. Ontological analysis bears, in particular, on ontological categories – such as those of quality and society – and ontological principles – such as the hypothesis that every concrete thing is in flux. The analysis we expect from scientific ontology concerns, in particular but not exclusively, the ontological categories and hypotheses that occur, either in a heuristic or in a constitutive capacity, in scientific research. Some such categories are those of thing, property, fact, and value. As for the ontological principles inherent in science, suffice it to mention the assumption that a society, far from being either an amorphous set of individuals, or a totality transcending individuals, is a system of interacting persons. More in Secs. 7 and 8.

However, the ontological analyst need not confine himself to listing and classing the ontological categories and principles actually employed – usually with neither acknowledgement nor apology – in scientific inquiry. He may be more ambitious and criticize certain ontological theses, be they proposed by philosophers or by scientists, and he may even propose his own theses on the matter. Take for example the generic or ontological notion of a mental event. The ontological analyst may not only examine the various construals of the expression 'mental event' but also have an axe to grind – or ought to if he is a philosopher at all. But he may not be able to accomplish much without the help of some general (ontological) theory of events (of all kinds) and some general (ontological) theory, or at least framework, of the mental. That is, ontological analysis, if profound, calls for ontological synthesis or theory, or at least a framework for such.

Ontological theories are of course those involving ontological categories. A general theory of changes of any kind, physical or mental, quantitative or qualitative, is an ontological theory. And a theory of mental events of any kind – feeling visceral processes, perceiving physical stimuli, forming images or concepts, etc. – is an ontological theory too, though a regional one instead of a universal theory. A complete ontology should include both *universal* and *regional* ontological theories. The former serve as frameworks for the latter, which will in turn illustrate and in a way test the former. Take again the case of mental events. In our own ontology the dualistic tenet concerning the mind-body problem makes hardly any sense because it presupposes that mental events are exceptional in that, unlike all others, they fail to be changes in some concrete thing. In fact in our ontology all and only things change, and every change (event of process) is a modification in the state of some concrete thing or other. Once this general ontological view has been adopted and systematized (made into a theory or at least a framework) there is no room left for changes that fail to be changes in or of some concrete thing, be it an electron or a neuron, a brain or a society. Only one view of mental events is consistent with that general ontological theory of change, namely that according to which the mind is a certain activity of the brain (Hebb, 1949). Moreover it is clear that any alternative view should be shown to be consistent with some general theory of change. This is certainly possible: a mentalistic (e.g. idealistic) ontology will include a mentalistic view of the physical. But of course such an ontology will be at variance with physics, chemistry and biology, none of which is animistic. We shall discuss these matters in detail in Vol. 4, Ch. 10. The point of bringing them up here was to defend the view that a full fledged ontology must consist of theories – or at least theory embryos – some universal and some regional, and that these theories should be mutually consistent. There is no salvation outside a system.

Needless to say, when using the word 'theory' we mean a hypothetical-deductive system rather than a stray opinion or an unsystematic set of opinions. In particular, an *ontological theory* is a theory that contains and interrelates ontological categories, or generic concepts representing components or features of the world. (Our own set of categories includes those of thing, property, law, possibility, change, spacetime, life, mind, value, and society.) Ideally, an ontological system or theory is a system, not just a set, of interrelated ontological categories. But we shall not always succeed in building theories proper. Sometimes we shall offer just ontological frameworks, which are constructs intermediate in structure between shapeless views and closed hypothetical-deductive systems. (For the general concept of a framework see Vol. 1, Ch. 2, Sec. 3.4, Def. 2.10.)

We shall agree to call an ordered triple $\mathbb{C} = \langle S, \mathbb{P}, D \rangle$ an *ontological framework* iff S is a set of statements in which occur only the predicate constants in the predicate family $\mathbb{P}$, which includes a nonempty set $\mathbb{O}$ of basic ontological concepts (i.e. categories), and the reference class of every P in $\mathbb{P}$ is included in the universe or domain D of hypothesized entities, or objects assumed to exist (in our case concrete objects or things). Though no substitute for an ontological theory, an ontological framework serves as a matrix for any number of ontological systems or theories, and has thus a guiding or heuristic power. (For the difference between theory and framework, and the heuristic role of the latter, see Bunge (1974e).) Indeed the mere display of D's and their properties (i.e. the members of $\mathbb{P}$) invites the formulation of hypotheses (statements in S) concerning the mode of being and becoming of those alleged entities.

An ontological framework will be *monistic* or *pluralistic* according as both the domain D of individuals and the collection $\mathbb{P}$ of their properties (and relations) are taken to be a single genus each or the union of two or more genera. (In short, monism is reductionistic in some sense or other, pluralism is not.) And, whether monistic or pluralistic, an ontological framework can be either *spiritualistic* or *naturalistic*, according as it postulates that D includes, or fails to include, a set of spiritual (disembodied) entities such as immaterial souls. But whereas according to spiritualistic monism all entities are spiritual and of the same kind (e.g. sensations, or ideas), according to pluralistic spiritualism all entities are spiritual but there are several types of them and these kinds are not reducible to one basic kind. And whereas according to monistic naturalism all entities are concrete and of the same kind (e.g. physical), pluralistic monism asserts that, although all entities are concrete, there are several irreducible kinds of them. The ontological frameworks and theories to be developed in the present volume and its sequel, Vol. 4, will turn out to be both *naturalistic and pluralistic*: we will assume only concrete existents but will assert their qualitative variety.

In summary, the goals of scientific ontology are *to analyze and to systematize the ontological categories and hypotheses germane to science.* These are the aims served by the construction of frameworks and theories close to science rather than closed to science.

6. Ontology and formal science

Formal science, or at least some of it, constitutes both the language and the formal skeleton of scientific ontology. In particular, scientific ontology presupposes *abstract mathematics*, including deductive logic. The choice of abstract mathematics as the formal backbone of metaphysics is only natural, for only abstract systems can be assigned alternative interpretations, in particular ontological interpretations. The theories of semigroups and of lattices, which are abstract, have far more potential use for metaphysics, and for philosophy in general, than number theory and the infinitesimal calculus, which are fully interpreted (in mathematical terms), hence nonabstract. In any case, metaphysics can make use of mathematics and must employ it if it is to become exact and contiguous with science. Is there more to it? Let us see.

Many philosophers, from Parmenides through Leibniz and Hegel to Gonseth and Scholz, have believed that logic is a sort of universal physics, i.e. the most general theory of being or becoming. The Leibnizians put it this way: (*a*) the real world is merely one among many equally possible worlds; (*b*) logic is true of all possible worlds, hence also of ours. (See e.g. Lewis, 1923; Scholz, 1941; Hasenjäger, 1966.) The rationale for this view is that a tautology of ordinary logic, such as $\ulcorner (x)(Px \vee \neg Px) \urcorner$, is true of anything, i.e. no restrictions are placed on the individual variable or the predicate variable. This is certainly true: logic refers to everything and holds for anything in any respect. It would seem therefore that logic is a sort of minimal ontology, a *physique de l'objet quelconque* (Gonseth, 1938).

However, although logic is indeed indifferent to the kind of referent, it does not describe or represent, let alone explain or predict, any factual items. The statement "Every organism is either alive or dead" concerns organisms but does not say anything definite about them; hence it is not a biological statement or even an ontological one. Rather, the proposition says something about the particular combination of "or" and "not" that occurs in it. Logic is the set of theories describing the properties of logical concepts – the connectives, the quantifiers, and the entailment relation. So much so that none of these concepts has a real counterpart or at least a unique one. Indeed, there are neither negative nor alternative things, properties, states, or events. Generalization, whether existential or universal, is strictly a conceptual operation. And the entailment relation has no ontic correlate either, although it has sometimes been likened to the causal relation. Further, if it be admitted that tautologies are analytic, hence a priori, they cannot be also synthetic a priori truths. In sum logic is not ontological.

The relation between logic and ontology is that of presupposition or logical priority: any cogent metaphysics presupposes logic. (This may be why Heidegger (1953) pushed logic aside with contempt and hatred.) In our work we shall take second order ordinary logic for granted. *Ordinary* logic because it is the one employed in mathematics, hence in theoretical science, neither of which has found use for any deviant logics. And *second order* logic because we want to analyze ontological statements such as "Everything has some property or other", and "Every property of a thing is related to some other properties of it".

The relation between ontology and mathematics is similar: if exact, ontology presupposes mathematics, but the latter is in turn free from any ontological commitments. Surely certain mathematical structures can be assigned ontic interpretations. For example group theory can be construed as describing all possible reversible transformations. But such interpretations, like those inherent in theoretical physics, go beyond pure mathematics. The latter stands on its own feet. If anyone wants to make use of it he is welcome: mathematics (including logic) is a service discipline. If a given field of research finds no use for mathematics, then this suggests only that it is in a backward state – that it has few clear notions and general propositions. The possibility of mathematizing a field of knowledge does not depend upon the subject matter but upon the state of the art, for what get mathematized are not facts but our ideas about them. As long as the latter are confused they are impregnable to mathematization, as was the case with the much overrated *impetus* concept (Koyré, 1957), later on with that of energy, and nowadays with that of the "neural correlate" of a mental event. So much for the relations between ontology and mathematics.

The relation between semantics and metaphysics is another matter. Tarski (1944, p. 363) held that ontology "has hardly any connections with semantics." However, the correspondence theory of truth presupposes that there is a real world – and this is a metaphysical assumption. Moreover the applications of any theory of reference calls for definite assumptions about the furniture of the world. Consider e.g. "The plane is 1 hr late". Does it refer to the plane, to both the plane and the airport, or only to the event of its arrival in the airport, or to both and the duration of the delay? The answer to this question depends on one's cosmology – on the assumed furniture of the world. If things are regarded as the only constituents of reality, the answer is that the statement concerns the plane and the airport; if events are declared the stuff of reality, then the referent will be the plane's arrival; and if time is assumed to have an absolute existence then it, too, will count among the referents. In sum semantics, or at least its application, does have metaphysical presuppositions. (More in Bunge, 1974c, and Vol. 2, Ch. 10, Sec. 4.3.) On the other hand ontological theories seem to have no semantic presuppositions although they do employ semantic concepts such as those of designation, reference, and representation.

To sum up. Deductive logic and pure mathematics, in particular abstract mathematical theories, are ontologically neutral. Precisely for this reason they can be used in building ontological theories. There is no a priori limitation on the variety of mathematical theories that can be employed in metaphysical research. The choice will depend largely upon the metaphysician's background and preferences. Some (e.g. Scholz, 1941) will try to make do with formal logic and model theory only. Others would like to extract an entire cosmology out of modal logic (see e.g. Munitz (ed.), 1971 and 1973). Suppes (1974) seeks instead "to replace the concept of logical empiricism by probabilistic empiricism" in constructing a probabilistic metaphysics. Finally, other philosophers will adopt what may be termed *mathematical opportunism*, or the strategy of using whatever mathematical theories seem promising. The present work exemplifies this strategy.

So much for the relation of metaphysics to formal science. (More in Bunge 1974c.) Let us now look at the other component of science.

7. THE ONTOLOGY OF SCIENCE

Metaphysics has often been contrasted with science for allegedly being speculative rather than empirical, hence irrefutable. So it is in many cases. However this is far from necessary: ontology can be consonant with science and just as scientific as physics, even though there will never be any metaphysical laboratories. Moreover we shall argue that (*a*) scientific research is guided or misguided by metaphysical principles – some good, others bad; (*b*) both basic science and technology have produced theories that are scientific as well as metaphysical, and (*c*) it is possible to build systems of scientific ontology. If this is so then there need not be any hostility between science and (scientific) metaphysics. There is not even a gap, let alone an abyss, between them: *ontology is general science* and *the factual sciences are special metaphysics*. In other words, both science and ontology inquire into the nature of things but, whereas science does it in detail and thus produces theories that are open to empirical scrutiny, metaphysics is extremely general and can be checked solely by its coherence with science.

That scientific research proceeds on a number of metaphysical hypotheses has been pointed out from time to time – e.g. by Woodger (1929, p. 228), Collingwood (1940), Russell (1948, pp. 506 ff.), Margenau (1941, 1950), Bunge (1961, 1967a I pp. 291 ff.), Harré (1961, 1972), Agassi (1964, 1975), Lakatos (1969), Harvey (1969), Körner (1970) and Rosenblueth (1970, Chs. 7 and 10). The following list of ontological principles occurring in scientific research (Bunge, 1974d) must suffice here:

M1 There is a world external to the cognitive subject. If there were no such world it would not be subject to scientific inquiry. Rather we would resort to introspection or to pure mathematics instead of attempting to discover the unknown beyond the self.

M2 The world is composed of things. Consequently the sciences of reality (natural or social) study things, their properties and changes. If there were real objects other than things it would be impossible to act upon them with the help of other things.

M3 Forms are properties of things. There are no Platonic Forms in themselves flying above concrete things. This is why (*a*) we study and modify properties by examining things and forcing them to change, and (*b*) properties are represented by predicates (e.g. functions) defined on domains that are, at least in part, sets of concrete objects. (Think of fertility, defined on the set of organisms.)

M4 Things are grouped into systems or aggregates of interacting components. There is no thing that fails to be a part of at least one system. There are no independent things: the borders we trace between entities are often imaginary. What there really is, are systems – physical, chemical, living, or social.

M5 Every system, except the universe, interacts with other systems in certain respects and is isolated from other systems in other respects. Totally isolated things would be unknowable. And if there were no relative isolation we would be forced to know the whole before knowing any of its parts.

M6 Every thing changes. Even the so-called ultimate components of matter end up by changing in the course of their transactions with other things. Even the supposedly stable particles can be absorbed by other systems or merge with their corresponding antiparticles to form photons that may in turn be absorbed.

M7 Nothing comes out of nothing and no thing reduces to nothingness. If this were not so we would make no effort to discover both the origin of new things and the traces left by things that have been destroyed.

M8 Every thing abides by laws. Whether natural or social, laws are invariant relations among properties, and they are just as objective as properties. Moreover a law is a property. If there were no laws we would never discover any and would not utilize them in order to explain, foresee and do. In particular the experimental method would not be feasible, for its essence is the deliberate and controlled wiggling of a component or a variable of some system to find out what effect that change may have on some other feature of the system. Our unrelenting attempt to find out facts of this kind presupposes that there are lawful relations among the items concerned.

M9 There are several kinds of law (nomological pluralism). There are causal (or predominantly causal) laws and stochastic laws, as well as laws exhibiting these two and even further modes of becoming. There are same-level (e.g. biological) laws and cross-level (e.g. psychosocial) laws.

M10 There are several levels of organization: physical, chemical, biological, social, technological, etc. The so-called higher levels emerge from other levels in the course of processes; but, once formed – with laws of their own – they enjoy a certain stability. Otherwise we would know nothing about organisms and societies before having exhausted physics and chemistry – which are inexhaustible anyway.

There are certainly further ontological principles utilized more or less overtly in the selection of problems, in the formation of concepts and hypotheses, in the design of techniques, and in the evaluation of results of scientific research. However, the ones we have listed are enough to establish the thesis that *all science presupposes some metaphysics*. Surely most scientists are not aware of this fact or else they call such principles ‘scientific’ (e.g. Rosenblueth, 1970). But the fact is there and once in a while we come across some great scientist who does realize that he is upholding certain metaphysical hypotheses, i.e. who has a definite and conscious world view that guides his research. Galilei, Descartes, Leibniz, Newton, Euler, d'Alembert, Priestley, Faraday, Darwin, Maxwell, Einstein, Born, Heisenberg, Schrödinger, J. B. S. Haldane, Bernal, and Dobzhansky – to cite but a few – were among the great scientists aware of some of their own ontological principles. And if scientific research is guided – or misguided – by ontological axioms then it behoves the historian of science to dig them up, and the philosopher of science to formulate them clearly, to justify or criticize them, and eventually to systematize them. This is in fact the task of the *metaphysics of science* (Bunge, 1974d, 1974f). Some of it will motivate, and some will be incorporated into, our system of scientific metaphysics.

To be sure not all of the metaphysical hypotheses proposed or adopted by scientists are reasonable or fertile: some are unreasonable or obstructive. To cite but a few examples: (*a*) the Copenhagen interpretation of quantum mechanics claims that, whereas macrophysical objects are objectively real, their atomic components come into being only as a result of observations; (*b*) the many-worlds interpretation of quantum mechanics holds that at each instant the world we live in branches off into countless parallel universes, mutually disconnected and therefore unknowable; (*c*) a number of scientists deny that chance is an objective mode of being and are therefore committed to the subjectivist or personalistic interpretation of probability: they are die-hard classical determinists; (*d*) according to a widespread interpretation of the theory of evolution, natural selection leads to gradual perfection, which in turn is what lends "meaning" to life; (*e*) behaviorism rests on the Aristotelian principle that nothing can change spontaneously: that the organism is a deterministic automaton every output of which is caused by some environmental input; (*f*) many eminent neurophysiologists and neurosurgeons believe that the mind is something separate from the body that does the minding; (*g*) while some social scientists claim that there are no societal properties and laws above those characterizing the individuals, others refuse to recognize that societal emergents are rooted to the individuals and their interplay. In sum, science is shot through with metaphysics – good and bad, fertile and sterile.

Let us now turn to some areas of scientific research where metaphysics makes itself felt with particular acuteness.

8. ONTOLOGICAL INPUTS AND OUTPUTS OF SCIENCE AND TECHNOLOGY

We shall presently defend the thesis that ontology can be found not only among the heuristic crutches that guide scientific research, but also (*a*) in some of the very *problems* handled by research, (*b*) in the *axiomatic reconstruction* of scientific theories, and (*c*) among *extremely general theories* in both basic and applied science.

That some of the most interesting scientific problems are at the same time metaphysical, is shown by the following examples:

(i) Is there an ultimate matter (*ápeiron* or *Urmaterie*)? This question triggered Heisenberg's 1956 theory of elementary particles.

(ii) Is life characterized by an immaterial principle (spirit, entelechy, *élan vital*, or any other substitute for ignorance), or has it emerged in a strictly natural evolutionary process? This question has prompted the current experimental studies on the origin of organisms.

(iii) Are biological species embodiments of Platonic archetypes, or just concrete populations, or something else? This question is being asked every day by almost every thoughtful taxonomist.

(iv) Is the mental anything disjoint from and beyond the neural? Are the two correlated, or do they interact, or finally is the mental a function of the nervous system? The latter alternative is no less than the motor of physiological psychology.

(v) Is a society anything beyond and above the individuals that compose it or are there societal laws in addition to the laws concerning individual behavior? This is one of the central disputes in today's methodology and philosophy of the social sciences.

The above questions have proved fruitful: they have triggered entire research lines. Alongside them there are other questions that result from a dubious metaphysics often at work in science, one that is a source of confusion and a sink of efforts. Here is a sample:

(vi) Why are μ-mesons different from electrons? (Why should they not be?)

(vii) Why are there no superluminal particles? (Why should there be?)

(viii) Is it possible to observe time reversal? (Inverting the sign of t in an equation of motion has no "operational" counterpart.)

(ix) Why should the study of behavior be hitched to an investigation of the behaving thing, i.e. the organism? (Because there is no change apart from a changing thing.)

(x) What is the liaison officer between body and mind? (Why presuppose that body and mind are separate entities?)

In sum, ontology inspires or blocks the most interesting research problems and plans. (This being so, the intervention of ontological principles in the design of research policies ought to be studied.)

A second area where ontology overlaps with science is the axiomatic foundation of scientific theories. If a scientific theory is axiomatized, some of the following concepts are likely to occur in it in an explicit fashion: part, juxtaposition, property, possibility, composition, state function, state, event, process, space, time, life, mind, and society. However, the specific axioms of the theory will usually not tell us anything about such fundamental and generic concepts: science just borrows them leaving them in an intuitive or presystematic state. Only ontology is interested in explicating and systematizing concepts which, since they are used by many sciences, are claimed by none. For example, physics asks not 'What is time?', biology 'What is life?', psychology 'What is the mental?', and sociology 'What is sociality?'. It is the task of ontology, jointly with the foundations of science, to try and supply answers to such questions and, in general, to clarify whatever idea science takes for granted or leaves in the twilight. That is, the metaphysician must fill some of the gaps in science.

There are such gaps in scientific knowledge not because certain facts are impregnable to the scientific method and require an esoteric intuition but because there are certain ideas which scientists use freely without bothering to examine, perhaps because they look obvious. The ontological theories that clarify and articulate the general ideas underlying a scientific theory T may be called the *ontological background* of T. This background becomes part and parcel of the axiomatized scientific theory (Bunge, 1967b, 1973b). That is, the metaphysical hypotheses contained in that background become blended with the scientific hypotheses proper. Note that such metaphysical ideas are now *constitutive* not just *regulative* or heuristic like those listed in Sec. 7. Whereas the latter may go down with the scaffolding the former remain in the theory.

In short, the axiomatic reconstruction of any fundamental scientific theory is bound to dig up, exactify and systematize certain ontological concepts. In this way the line between ontology and science is erased.

A third region where science merges with ontology is constituted by certain extremely general theories such as Lagrangian dynamics, the classical theory of fields (of any kind), and the quantum theory of fields (misnamed 'axiomatic field theory'). All these theories are generic in the sense that, far from representing narrow species of things, they describe the basic traits of whole genera of things. For example Lagrangian dynamics, which started out as a reformulation of classical particle mechanics, was eventually generalized to cover mechanical, electrical, and even biological and social systems. As long as the Lagrangian function is not specified, and its independent variables are not interpreted, the theory remains so general as to be indistinguishable from metaphysics. In fact Lagrangian dynamics may have been the earliest member of scientific metaphysics.

In sum, science itself has produced ontological theories by a process of generalization. Here again there is no border to be drawn.

A fourth and last area where ontology mingles with science is contemporary technology. In fact some of the theories included in the so-called information sciences and in systems theory are so general, and at the same time so precise, that they qualify as theories in scientific metaphysics. For example a general control (or cybernetic) theory will apply not only to machines with feedback loops but also to goal-seeking systems (members of higher animal species). Being so general, a theory of this kind will not represent any details of the system of interest: for example it will be insensitive to the nature of the materials and mechanisms involved. Hence it will not replace any of the special theories built by scientists and technologists. But such general theories provide insights and they link previously isolated fields of research. (Cf. Bunge, 1977a.)

In short, post-war technology has produced a number of ontological theories. In fact it has been more fertile than academic metaphysics.

The upshot of our discussion so far is this: science, whether basic or applied, is permeated with ontological ideas, now heuristic, now constitutive – so much so that some scientific theories are at the same time metaphysical. More precisely: every scientific theory, if extremely general, *is* ontological; and every ontological theory, if exact and in tune with science, *is* scientific. Therefore the expression *scientific metaphysics*, though perhaps shocking at first, designates an existing field and one that is hardly distinguishable from science.

True, ontological theories, whatever their degree of scientificity, cannot be tested empirically. Nor, for that matter, can extremely general scientific theories. (For example the basic equations of continuum mechanics cannot even be solved, let alone put to the test, unless they are adjoined special assumptions concerning the forces, the mass distribution, and the internal stresses.) Like the generic theories of science, those of ontology are *vicariously testable*, i.e. through the checking of more special theories gotten from the general ones by conjoining them with subsidiary assumptions (Bunge, 1973a, Ch. 2). In sum, concerning testability as much as concerning reference, there is no distinction between an ontological theory and an extremely general scientific one.

Because there is no clear-cut border between science and ontology, it is pointless to try to find the correct criterion for the strict demarcation between these two fields. The demarcation problem, which has occupied so many eminent brains from Kant on (see e.g. Popper, 1963), has disappeared. Another, far more interesting and rewarding task, has replaced it: namely that of digging up the metaphysics of science and technology, and building a scientific ontology.

9. USES OF ONTOLOGY

Metaphysics is a traditional branch of philosophy and as such it need not be justified in the eyes of the philosopher provided he is either pre-Kantian or post-positivist. (It has been remarked many times, and rightly so, that an antimetaphysician is one who holds primitive and unexamined metaphysical beliefs.) A philosopher may dislike the word and prefer some substitute for it, such as 'ontology', 'philosophical cosmology', 'general science', or 'scientific outlook'. But, since he acknowledges the existence of certain problems that have traditionally been called 'metaphysical', unless he is an irrationalist he won't deny that those problems deserve an answer and, moreover, that the best answer to any theoretical problem is supplied by a theory.

If metaphysics is a part of philosophy then it must be related to the other branches of this field: otherwise it would not constitute such a part. It can certainly be distinguished from the other branches of philosophy – mainly logic, semantics, epistemology, value theory, and ethics – but not separated from them. Thus it makes a lot of difference for ontology if one adopts a realist epistemology or a phenomenalist one: with the former one will attempt to disclose the structure of reality, with the latter only the structure of appearance. In turn, one's epistemology depends to a large extent upon one's ontology: if the latter is naturalist one will conceive of epistemology as the ontology of knowing; but if one adopts a spiritualist ontology then the way to a subjectivist epistemology is open. To generalize: ontology is closely linked to the rest of philosophy.

However, the close interactions between ontology and its sister disciplines should not blind us to their differences. In particular we must not mistake the categories of knowing (e.g. truth) with those of being (e.g. property). And yet this distinction is sometimes overlooked. For example Strawson (1959) has claimed that material bodies are basic among particulars, not because they are the ultimate constituents of everything, but because every particular is allegedly identified or reidentified in terms of material bodies. Surely material bodies are a condition for our knowledge of the external world. Yet so are our senses and our intellect and yet these are not basic particulars. In sum, while defending the unity of philosophy, i.e. the interrelations among its branches, we must not forget their differences.

In any case ontology hardly needs a defense among contemporary philosophers. It has even become fashionable among them to talk about ontology. On the other hand scientists will not be impressed by the statement that metaphysics is worthy because it happens to be part and parcel of philosophy. (After all, they were told by philosophers, over a century ago, that metaphysics is balderdash.) But scientists might learn to tolerate ontology if shown that, though at first blush useless (or even nonsense), it can be of some use, namely in the following ways:

(i) *Metaphysics can help dissolve pseudoquestions* that arise in science and originate in misconceptions. Example: "What is the root of the arrow of time: irreversibility, ignorance of microscopic details, experiment, the expansion of the universe, or what?" This question dissolves upon realizing that time has no objective arrow: that only processes can have an "arrow", i.e. be irreversible. But this realization requires a full-fledged theory of time, and this is a typical task for ontology.

(ii) *Metaphysics can dig up, clarify, and systematize* some basic concepts and principles occurring in the course of scientific research and even in scientific theories – constructs that are common to a number of sciences, so that no single science takes the trouble to regiment them, as is the case with "property" and "space". Some of these constructs are included in the ontological background of scientific theories: recall Sec. 8.

(iii) *Metaphysics can suggest new scientific problems, research plans, and even theories.* Example: the mechanistic metaphysics born in the early 17th century inspired scientific research for two centuries. Another: Nobel laureate Yukawa (1973) blames the current cult of blind computation and measurement, and the concurrent lack of a deep understanding of high energy physics, on the lack of an inspiring metaphysics.

(iv) *Metaphysics can render service by analyzing fashionable but obscure notions* such as those of system, hierarchy, structure, event, information, and possible world, as well as by criticizing popular misconceptions such as that the world is a collection of facts and the mind the control of the body.

(v) *Metaphysics can perform another public service by examining the ontological kernel of the current ideologies* and finding out whether they live up to the standards of contemporary intellectual work.

Briefly, ontology is far from useless: sometimes it fuels good work, at other times it is an obstacle to research, and at all times it is there, in the very midst of our speculations about the world.

10. Concluding remarks

By way of summary:

(i) Ontology can be exact or inexact, according as it uses any formal tools or fails to employ them.

(ii) Exact (or mathematical) ontology may be meaty or hollow, according as it tackles or evades the great problems concerning the nature of reality.

(iii) Exact and meaty metaphysics may or may not be scientific, according as it is contiguous with science or alien to it.

(iv) Scientific ontology is a collection of general or cross-disciplinary frameworks and theories, with a factual reference, mathematical in form and compatible with – as well as relevant to – the science of the day.

(v) There is no gap between good metaphysics and deep science: every scientific theory with a wide scope may be regarded as metaphysical, and every ontological theory that brings together and generalizes scientific results, or else occurs in the background of an axiomatized scientific theory, qualifies as scientific.

(vi) Science, both basic and applied, teems with metaphysical concepts and hypotheses: it presupposes certain ontological principles of both the heuristic and the constitutive kind, and it is a powerful source of metaphysical conjectures. In fact some theories are both metaphysical and scientific. Since there is no definite border line between the two fields, there is no demarcation problem left.

Enough of meta-metaphysical discourse. Let us build a new ontology.

CHAPTER 1

SUBSTANCE

One of Descartes' *regulae philosophandi* counsels us to start by simplifying, and one of our own maxims (Introduction, Sec. 4) enjoins us to hypothesize unobservables in order to account for appearances. A radical joint implementation of these two rules consists in stripping real things of all their properties – only for a start of course. What remains is the qualitatively indeterminate particular, the *bare individual.* Our bare individual will turn out to be similar to, though not identical with, Plato's formless matter and Aristotle's primary substance. Similar because neither has inherent properties (except possibly composition in our case), different because our bare individuals are endowed with the capacity of associating, i.e. of forming composite entities. Whereas each bare individual is nondescript, an aggregate of bare individuals has the definite property of being composed, i.e. of having parts. The association of bare individuals is then a beginning of complexity and thus a step towards realism.

Surely every real thing consists of some definite stuff endowed with definite properties: there are no bare individuals except in our imagination. This is a tacit assumption of factual science. In other words the Platonic concept of formless matter and the Aristotelian notion of unchanging substratum have been shelved by science. Still, although we no longer believe in the actual existence of nondescript things, we find it methodologically convenient to feign them – only to endow them with further properties later on, namely in Ch. 2. A fully qualitied individual, if substantial or concrete, will be called a *thing* in Ch. 3. And a complex thing with coupled components will be termed a *system* in Ch. 5. This is how we shall tackle the basic problem of traditional metaphysics, i.e. that of substance and attribute. First, then, we address ourselves to the concept of substance.

1. Association

1.1. *Concatenation and Its Ontological Interpretation*

Individuals can associate to form further individuals. When neither the kind of individual nor the manner of association are specified, we have the (bare) association of (bare) individuals. This notion of association can be formalized by the algebraic concept of concatenation, which is elucidated in the theory of semigroups. (Cf. Ljapin, 1963.)

A semigroup is a set S together with a binary, internal and associative operation $\circ$ of concatenation. The operation is said to be internal because it is defined in S, i.e. because S is closed under $\circ$. (I.e., concatenation does not give rise to individuals alien to S.) And concatenation is associative because it satisfies the law of associativity: if $x, y, z \in S$, then $x \circ (y \circ z) = (x \circ y) \circ z$. More succinctly: the structure $\langle S, \circ \rangle$, where S is a nonempty arbitrary set and $\circ$ a binary operation in S, is called a *semigroup* if and only if (i) S is closed under $\circ$, and (ii) $\circ$ is associative in S. The semigroup $\langle S, \circ \rangle$ is finite if S has a finite number of elements, infinite otherwise. A simple example of a finite semigroup is this: $S = \{0, 1, 2\}$ with the multiplication (concatenation) table

$\circ$	0	1	2
0	0	1	2
1	1	1	2
2	2	2	2

And a simple example of an infinite semigroup is the set $\mathbb{N}$ of positive integers with respect to addition. In other words the structure $\langle \mathbb{N}, + \rangle$ is a model of a semigroup. (For the semantic concept of a model see e.g. Vol. 2, Ch. 6, Sec. 2.)

If we interpret S as the totality of bare individuals and $\circ$ as their pairwise association, we obtain the simplest, most basic and most useful theory in scientific metaphysics: we call it *association theory*.

We shall do slightly better by adding the assumption that our semigroup, when written additively, contains the null element, and interpret the latter as the *null individual*. Roughly, the neutral element $\Box$ of S is that (fictitious) individual which, when coupled to an arbitrary member x of S, leaves x unaltered, i.e. $x \circ \Box = \Box \circ x = x$. Formally, $\Box$ is the two-sided zero of S. The new structure, $\langle S, \circ, \Box \rangle$, is called an (additive) semigroup with neutral element, or *monoid* for short. This structure is slightly richer than the corresponding semigroup, for it is more specific: now one of the elements of S, namely $\Box$, is no longer a faceless individual but is distinct from all the others – not by an intrinsic quality but by its inane behavior in company. Since $\Box$ is the nonentity or nonbeing, it is even more fictitious than any of the nondesignated elements of S. The introduction of $\square$ is justified by the fact that it will allow us to state that nothing starts from $\square$ or ends up in $\square$ – an important ontological principle indeed. (More on $\square$ in Sec. 3.1.)

1.2. *Axiomatic Foundation of Association Theory*

We postulate then that the set of individual things, when stripped of their properties and interactions – except for the property of composition – has the monoid structure. Moreover we assume that the association order makes no difference, i.e. that the concatenation $\circ$ is *commutative*: if $x, y \in S$, then $x \circ y = y \circ x$. We also assume that the association of an individual with itself leaves it unchanged, i.e. that the elements of S are *idempotent*: $x \circ x = x$. (On the other hand we do not postulate that the association operation has the cancellation property, i.e. that, for all a, b, x in S, the equality $a \circ x = b \circ x$ implies that $a = b$.) The previous assumptions entail that the result of the association of two individuals is non-null provided at least one of them is not null: for all x, y in S, if $x \neq \square$ or $y \neq \square$, then $x \circ y \neq \square$. (I.e., association does not result in annihilation.) We call this property *substance conservation*.

In sum, we assume that the set S of bare individuals is a commutative monoid of idempotents. And we assume that this structure carries over to all real things: that is, we postulate that the totality of concrete or substantial particulars, whether viewed as bare or as fully qualitied, has the monoid structure. (Caution: S is not the world, or reality, but the set of its elements. The world is one more individual, the supreme one, whereas a set is a concept. More on this in Sec. 3.1.) The preceding assumptions are compressed in the following compound

POSTULATE 1.1 Let S be a non-empty set, $\square$ a selected element of S, and $\circ$ a binary operation in S. Then the structure $\mathcal{S} = \langle S, \circ, \square \rangle$ satisfies the following conditions:
(i) $\mathcal{S}$ is a commutative monoid of idempotents;
(ii) S is the set of all substantial or concrete individuals;
(iii) the neutral element $\square$ is the null individual;
(iv) $\circ$ represents the association of individuals;
(v) the string $a_1 \circ a_2 \circ \ldots \circ a_n$, where $a_i \in S$ for $1 \leqslant i \leqslant n$, represents the individual composed of the individuals a_1 to a_n.

Remark 1 This axiom has a factual content insofar as it refers ultimately to real things and is far from being tautological. *Remark 2*

However, our postulate is hardly refutable. For, if any thing seemed to fail to satisfy our axiom, we would not say it associates. (More on this peculiarity of ontological theories in Bunge (1973a).) *Remark 3* The condition of commutativity is the only law statement involved in the above postulate. It can be rephrased thus: Association is a symmetric relation in the sense that "x associates with y" is equivalent to "y associates with x". *Remark 4* Since Postulate 1.1 states that entities can associate to form complex entities, it is inconsistent with Parmenides' block universe. *Remark 5* We have interpreted concatenation as association, whether natural or brought about by man. A narrower, operationist interpretation would view $\circ$ as the putting together of perceptible things by somebody. *Remark 6* It would be interesting to investigate even poorer structures, namely the nonassociative algebras such as groupoids (cf. Bruck, 1958). But we shall not take up this matter here. Likewise we shall not study the ontological applications of noncommutative semigroups.

Before introducing further assumptions we need some definitions. First

DEFINITION 1.1 An individual is *composite* iff it is composed of individuals other than itself and the null individual. I.e., $x \in S$ is composite iff there exist substantial individuals $y, z \in S$ such that $x = y \circ z$ and each differs from x as well as from $\square$. Otherwise the individual is *simple*.

Remark 1 According to the definition, the trivial composition $x \circ x = x$ does not count. *Remark 2* Note the difference between the composite entity $z = x \circ y$ and the set $\{x, y\}$ of its constituents or components. The latter is a concept and does not satisfy Postulate 1.1.

COROLLARY 1.1 The null individual is simple.

Proof $\square$ is simple because, for every non-null individual x, $x \circ \square \neq \square$ by Postulate 1.1 (ii).

Remark The null individual need not be the only simple one.

And now another key concept:

DEFINITION 1.2 If x and y are substantial individuals, then x is a *part* of y iff $x \circ y = y$. Symbol: $x \sqsubset y$.

Justification The relation $\sqsubset$ is a *partial order* relation, which is what the part-whole relation is supposed to be. In fact, from the previous definition and the properties of $\circ$ (Postulate 1.1) it follows that, for all x, $y, z \in S$,

(i) $\sqsubset$ is *reflexive*, i.e. $x \sqsubset x$;
(ii) $\sqsubset$ is *asymmetric*, i.e. if $x \neq y$ then: $x \sqsubset y \Rightarrow \neg(y \sqsubset x)$;
(iii) $\sqsubset$ is *transitive*: $x \sqsubset y \ \& \ y \sqsubset z \Rightarrow x \sqsubset z$.

Remark 1 The part-whole relation $\sqsubset$ must not be mistaken for the set inclusion relation $\subseteq$, much less for the set membership relation $\in$. Indeed whereas $\subset$ and $\in$ are relations among concepts of a certain kind (sets), $\sqsubset$ is a relation among substantial individuals. Besides, unlike $\sqsubset$, $\in$ is not transitive. (For example, if $A = \{x\}$ and $B = \{A\}$, then $x \in A$ and $A \in B$ but $x \notin B$.) *Remark* 2 We could have defined the converse $\sqsupset$ of $\sqsubset$. This is actually Whitehead's concept of "extension": x *extends over* y iff $x \circ y = x$ (Whitehead, 1919). But because here neither relation is construed in spatial terms, we might as well avoid geometric metaphors at this point. Indeed, an individual on our planet and another in a distant galaxy may be taken to associate to form a third individual, so that each component will be a part of the whole, just as much as the two components of a miscible fluid poured into a glass. *Remark 3* With the help of the previous definitions we can easily prove that

For any $x \in S: x$ is simple iff, for all $y \in S$, $y \sqsubset x \Rightarrow y = x$ or $y = \Box$.

Remark 4 The condition "For every $x \in S : x \circ \Box = \Box \circ x = x$", defining the null individual, can now be read as follows: The null individual is part of every individual. We proceed to introduce the dual of this concept:

POSTULATE 1.2 There exists an individual such that every other individual is part of it. I.e., $(\exists x)[x \in S \ \& \ (y)(y \in S \Rightarrow y \sqsubset x)]$.

DEFINITION 1.3 The universal individual introduced by Postulate 1.2 is called *the world* and is denoted by ▯.

Remark 1 Note again that the world, i.e. ▯, is an individual not to be confused with the set S of all individuals, which is a concept not a physical object. *Remark 2* The distinguishing algebraic property of ▯ is this: for every $x \in S$, $x \circ \text{▯} = \text{▯} \circ x = \text{▯}$. See Corollary 1.5.

DEFINITION 1.4 Let x denote an arbitrary object and ▯ the world. Then
(i) x is *worldly* iff $x \sqsubset \text{▯}$;
(ii) x is *unworldly* iff x is not worldly (i.e. if x does not satisfy Postulate 1.1 and a fortiori x is not a part of ▯).

Example The world is worldly but cosmology is not.

DEFINITION 1.5 Let x and y be substantial individuals. Then

(i) x and y are *detached* (*separate*) iff neither is part of the other:

$$x \rceil\!\lfloor y =_{df} \neg(x \sqsubset y) \,\&\, \neg(y \sqsubset x);$$

(ii) x and y *intersect* iff they are not detached:

$$x \bigcirc y =_{df} x \sqsubset y \vee y \sqsubset x.$$

The part-whole relation allows us to define yet another interesting concept that will be used throughout this work:

DEFINITION 1.6 The *composition* of an individual equals the set of its parts. I.e., let $\mathscr{C}: S \to 2^S$ be a function from individuals into sets of individuals, such that $\mathscr{C}(x) = \{y \in S \mid y \sqsubset x\}$ for any $x \in S$; then $\mathscr{C}(x)$ is called the *composition* of x.

Let us now derive a few consequences from our assumptions and definitions.

1.3. *Consequences*

First of all the trivial yet necessary

COROLLARY 1.2 The totality S of substantial individuals is partially ordered by the part-whole relation $\sqsubset$. I.e., $\langle S, \sqsubset\rangle$ is a poset.

Proof Immediate by recalling the justification of Definition 1.2.

THEOREM 1.1 The association of any two individuals is the supremum (least upper bound or l.u.b.) for them with respect to the part-whole ordering:

$$\text{If } x, y \in S \quad \text{then} \quad sup\{x, y\} = x \circ y.$$

Proof By associativity and idempotence $x \circ (x \circ y) = (x \circ x) \circ y = x \circ y$. By Definition 1.2 the last formula is the same as: $x \sqsubset (x \circ y)$. Likewise $y \circ (x \circ y) = x \circ y$, whence $y \sqsubset (x \circ y)$. Hence $x \circ y$ is an upper bound of x and y. It is also their least upper bound. In fact call z an upper bound of x and y, i.e. $x, y \sqsubset z$. Then $x \circ y \circ z = z$. Thus $x \circ y \sqsubset z$.

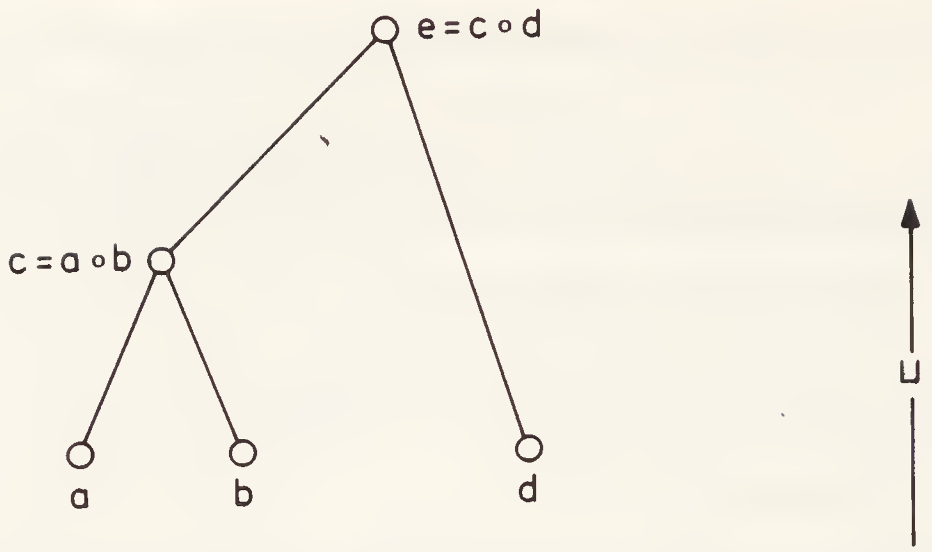


Fig. 1.1. All entities associate and they are ordered by the ⊏ relation.

The concept of supremum generalizes to arbitrary (in particular infinite) sets that of concatenation, which is defined only for pairs of individuals. We had postulated that, if x and y are substantial individuals, so is $x \circ y$ (Postulate 1.1). We now generalize that hypothesis to an arbitrary set of individuals: see Figure 1.1. That is, we assume

POSTULATE 1.3 Every set $T \subseteq S$ of substantial individuals has a supremum, which is denoted by $[T]$. I.e., for any $T \subseteq S$, there exists an individual $[T] \in S$ such that

(i) $x \sqsubset [T]$ for all $x \in T$;

(ii) if $y \in S$ is an upper bound of T, then $[T]$ precedes y: i.e., if $x \sqsubset y$ for all $x \in T$, then $sup\, T = [T] \sqsubset y$.

This axiom allows us to elucidate an important metaphysical notion:

DEFINITION 1.7 Let $T \subseteq S$ be a set of substantial individuals. Then the *aggregation* or *association* of T, or $[T]$ for short, is the supremum of T. I.e., $[T] = sup\, T$.

Remark 1 If T is a finite set then its aggregation $[T]$ is the association of all the members of T. I.e., if $T = \{x_1, x_2, \ldots, x_n\}$, then $[T] = sup\, T = x_1 \circ x_2 \circ \ldots \circ x_n$. In particular, the aggregation of a singleton is its sole occupant: if $T = \{x\}$, then $[T] = x$. *Remark 2* Because association is a finitary operation it is not possible to define the aggregation of an arbitrary (possibly infinite) set $T \subseteq S$ as the concatenation of all its members. This is why we had to employ the more general concept of *sup* or l.u.b. Let us now apply it to the entire collection of substantial individuals.

THEOREM 1.2 The world is the aggregation of all individuals:

$$\text{▯} = [S] = \mathit{sup}\, S.$$

Proof By Postulate 1.2, ▯ exists and is the last individual, i.e. for every $x \in S$, $x \sqsubset \text{▯}$. But this individual fits the conditions of Definition 1.7, i.e. $\text{▯} = \mathit{sup}\, S$.

The preceding material is compressed into

THEOREM 1.3 The ordered quadruple $\langle S, \circ, \square, \text{▯} \rangle$ is a sup-semilattice with least element □ and last element ▯ with respect to the part-whole relation $\sqsubset$.

Proof By Theorem 1.1 there exists a supremum for any two individuals, namely their association. Besides, □ is part of every individual, so it lies at the bottom of the net. Dually, ▯ contains every individual, so it perches on top. These conditions, together with the commutativity of association, define a sup-semilattice. See Figure 1.2.

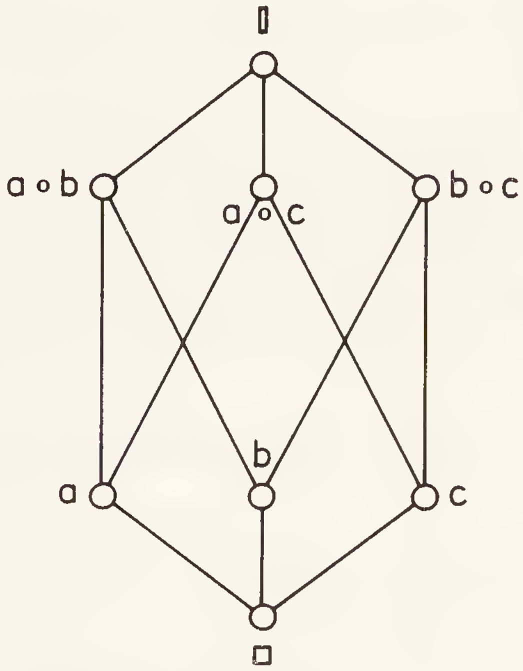


Fig. 1.2. Semilattice generated by $S = \{\square, a, b, c, \text{▯}\}$.

THEOREM 1.4 Every substantial individual is the aggregation of its parts, i.e. of its composition. I.e., for every $x \in S$,

$$x = [\mathscr{C}(x)] = \mathit{sup}\, \{y \in S \mid y \sqsubset x\}.$$

Proof Apply Definition 1.7 to $\mathscr{C}(x)$ characterized by Definition 1.6.

COROLLARY 1.3 Every individual with a finite composition is the association of its predecessors with respect to $\sqsubset$.

Example If $\mathscr{C}(c)=\{a, b\}$, then $c=a\circ b\circ c$, and $a, b\sqsubset c$.

Now two further consequences of Postulate 1.2. First

THEOREM 1.5 There is just one world.

Proof By Postulate 1.3 every $T\subseteq S$ has a supremum and by Theorem 1.2 $[S]=\Box$. And all suprema are unique. *Alternative proof* Suppose there were a world W different from $\Box$. Then since every substantial individual is part of $\Box$, so must be W. Hence W deserves being called a subworld rather than the world.

Remark Like many statements in this chapter, this one is mathematically trivial. However, it is philosophically nontrivial in view of the fashionable possible worlds metaphysics (Ch. 4, Sec. 6.2), the many-worlds (mis)interpretation of quantum mechanics (Ch. 4, Sec. 6.6), and Popper's three worlds.

COROLLARY 1.4 Whatever is unworldly is not a substantial individual.

Proof By Postulate 1.2, for every $x\in S$, $x\sqsubset\Box$. By contraposition, if $\neg(x\sqsubset\Box)$ then $\neg(x\in S)$.

Remark Classes, such as S, every one of its subsets, and its power set, are unworldly. So is any object constructed out of classes. So is any object that fails to satisfy the clauses of Postulate 1.1. Any object other than a substantial individual (or member of S) will be called a *construct*. I.e., nothing is both a construct and a substantial individual. More on this in Sec. 3.2.

Next we prove a bunch of conservation principles. First

THEOREM 1.6 If association ensues in the null individual then the associates were null to begin with:

$$\text{For all } x, y\in S,\ x\circ y=\Box\Rightarrow x=\Box\ \&\ y=\Box.$$

Proof Assume the antecedent, i.e. $x\circ y=\Box$. Now, $x=x\circ\Box=$ $=x\circ(x\circ y)=(x\circ x)\circ y=x\circ y=\Box$ by associativity, idempotence, and hypothesis. Similarly for y.

THEOREM 1.7 No substantial individual associates destructively with another: If $x\neq\Box$ then there is no $y\in S$ such that either $x\circ y=\Box$ or $y\circ x=\Box$.

Proof Assume $x \in S$ different from $\square$ and suppose there is in S an element y such that $x \circ y = \square$. Then, by Theorem 1.6 and modus tollens, $x = \square$, contrary to hypothesis.

Remark 1 In other words, except for $\square$, S has no inverses. Hence (*a*) S cannot possibly attain a group structure, and (*b*) the belief that there is an anti-thing for any given thing, such that they annihilate each other, is wrong. *Remark 2* It is sometimes said that, when particles of matter and antimatter join, they "annihilate" reciprocally. There is no such annihilation; there is rather the conversion of a pair of material entities (e.g. an electron and a positron) into a field chunk (e.g. a gamma ray photon).

COROLLARY 1.5 Provided it starts with something, association does not end up in nothingness:

$$\text{For all } x, y \in S, x \neq \square \;\&\; y \neq \square \Rightarrow x \circ y \neq \square.$$

Proof By Theorem 1.6 and contraposition.

COROLLARY 1.6 Nothing comes by association out of nought.

Proof This is the ontological interpretation of the implicit definition of the null individual, i.e. $x \circ \square = \square \circ x = x$.

COROLLARY 1.7 The world is not enlarged by associating it with any of its parts: For any $x \in S$, $x \circ \text{▯} = \text{▯} \circ x = \text{▯}$.

Proof By the definition of world.

Remark Corollaries 1.5 and 1.6 were first stated by Epicurus, then restated by Lucretius, in the form of postulates: "Matter is indestructible" and "Matter is uncreatable". (Cf. DeWitt, 1954; Lucretius 55 B.C.)

Another axiom of Greek atomism was that every thing is either basic (simple) or composed of basics (simples). We shall adopt this hypothesis that there are basic building blocks out of which every thing, from molecule to star, from cell to ecosystem, is composed, because it has been tremendously fruitful in modern science. That is, we lay down

POSTULATE 1.4 There is a proper subset B of S such that every substantial individual is the aggregation of members of B. I.e., for every $x \in S$ other than the null individual, there is a unique subset $B_x \subset B$ such that $x = [B_x]$.

Consequently instead of using the full aggregation function $[\ ]: 2^S \to S$, we need only its restriction to the power set of the simples or basics, i.e. the bijection $[\ ]|2^B: 2^B \overset{\sim}{\to} S$.

We shall make explicit use of Postulate 1.4 in our theory of space (Ch. 6). By the way, simples (basics) need not be spatially unextended. By Definition 1.1 a simple or basic individual is one that is not composed of any other individuals. For example, according to present day theories an electron is such a simple or basic because it is an indivisible whole even though both its electromagnetic field and its ψ-field are spatially extended. Another way of putting this is to say that electrons, though spatially extended, have no proper parts. In general,

$$\text{If} \quad x, y \in B \quad \text{and} \quad x \sqsubset y, \quad \text{then} \quad x = y.$$

So much for association theory. We shall generalize it in Sec. 2, but before that another two interesting notions have to be explored.

1.4. *Atom Aggregates*

A tacit assumption of association theory is that no two individuals are identical. In fact if a subset T of S consists of two identical individuals then, by the extensionality principle of set theory, T is a singleton. Association theory makes therefore no room for the possibility that some individuals come in multiple copies, as in the case of the letters of an alphabet, which can be duplicated *ad libitum*. Let us explore this alternative.

Assume there are simples and that every thing is either a simple or an association of simples – i.e. accept Postulate 1.4. And assume also that a single simple a may be found repeated any number of times, forming the composite entities $a \circ a$, $a \circ a \circ a$, etc. Two simples a and b will of course give rise to a greater variety: $a \circ a$, $a \circ a \circ a$, ..., $b \circ b$, $b \circ b \circ b$, ..., $a \circ b$, $a \circ a \circ b$, $a \circ a \circ b$, ..., $a \circ b \circ b$, $a \circ b \circ b \circ b$, ... Obviously even a few basic individuals, provided there are enough copies of them, will give rise to both numerosity and variety: this was a capital discovery of the ancient atomists, both Greek and Indian: *Multum ex parvo.*

To formalize the previous idea we start from a basic set A of distinct atoms or simples and give up the postulate of idempotency. Call then $A = \{a_i | i \in \mathbb{N}\}$ the set of basic elements or atoms and allow them to associate, as in $a_i \circ a_j$ and $a_j \circ a_i$, counting the association of two replicas of a given atom a_i, i.e. $a_i \circ a_i$, as a third individual distinct from a_i. Since an atom aggregate is not an atom itself, A is not closed under association. What is closed is the set A^* of all the concatenates. For example, a single atom generates infinitely many complexes. Indeed, if $A=\{a\}$, then $A^*=\{a^n | n \in \mathbb{N}\}$, where a^n abbreviates the concatenation of n identical copies of a. The set A^* is isomorphic to the set of natural numbers, so a universe formed of atoms of a single kind could still be infinitely complex. In general, if A is a non-empty set of atoms or basics, the collection A^* of concatenates of members of A is called the *free monoid generated by* A.

Example 1 The set of all conceivable chemical compounds is the free monoid generated by the set of chemical elements or atom species – just over 100. However, the set of compounds permitted by the laws of chemical binding is a finite subset of that free monoid. Thus the molecule HHHHHH, though thinkable, is not chemically possible. More on real possibility as distinct from conceptual possibility in Ch. 4.

Example 2 The minimal structure of a language with alphabet Σ is the free monoid generated by Σ: cf. Vol. 1, Definition 1.1.

We could assume, as a crude model of the world, that its furniture is a free monoid over some finite set S of simples. However, we know this to be an oversimplification: in the real world things do not come in strictly identical copies but, at most, in semi-copies. Moreover not all conceivable concatenations are really possible. However, certain subsets of S – elementary particles, atomic nuclei, atoms, molecules, cells, etc. – do come in *almost* identical copies. Hence for such things the free monoid over a finite set of atoms or simples is a sufficiently adequate (approximately true) model.

1.5. *Clustering*

The associations and dissociations of real things are brought about in specific ways – e.g. by motions or by forces –, hence their study pertains to the special sciences. However, certain features of such processes, such as the grouping of a collection of individuals into clusters, can be described a priori. So much so that this problem is the subject of a branch of pure mathematics, namely combinatorics (or combinatorial analysis). Let us take a glimpse at a couple of typical problems in combinatorics.

Problem 1 In how many ways can n individuals associate if association (concatenation) is not assumed to be commutative? That is, how many permutations of n objects (conceptual or physical) are possible? Answer: $n!$ (i.e. $1.2.3 \ldots n$). Example: the individuals a, b, c can associate thus: abc, bca, cab, bac, acb, cba.

Problem 2 Consider n objects grouped into clusters (or cells or blocks) of m individuals each (where $m \leqslant n$) without regard to order. How many such clusters are there? Answer: there are $n!/(n-m)!m!$ combinations of n individuals taken m at a time.

Problem 3 A set of n objects is partitioned into m clusters, with $m < n$. How many clusters will contain two or more individuals? Answer: at least one (Dirichlet's pigeonhole principle).

Problem 4 Consider a set of n objects. In how many ways can it be partitioned? For example, $S = \{a, b, c\}$ can be partitioned in the five following ways:

$$\{a\}, \{b\}, \{c\} \qquad \langle 3, 0, 0\rangle$$

$$\{a, b\}, \{c\}; \{b, c\}, \{a\}; \{a, c\}, \{b\} \qquad \langle 1, 1, 0\rangle$$

$$\{a, b, c\} \qquad \langle 0, 0, 1\rangle.$$

Every partition is characterized by its type $\langle m_1, m_2, \ldots, m_n\rangle$, where m_1 is the number of singles, m_2 that of couples, and so on. In every partition the number of blocks is $b = m_1 + m_2 + \ldots + m_n$, and $n = m_1 + 2m_2 + \ldots + nm_n$.

Like arithmetic, combinatorics is a theory of objects of any kind. It is therefore too general to pass for an ontological theory. (Recall the Introduction, Sec. 6.) But of course it can enter ontology, along with other mathematical theories, as an auxiliary.

1.6. *Historical Remark*

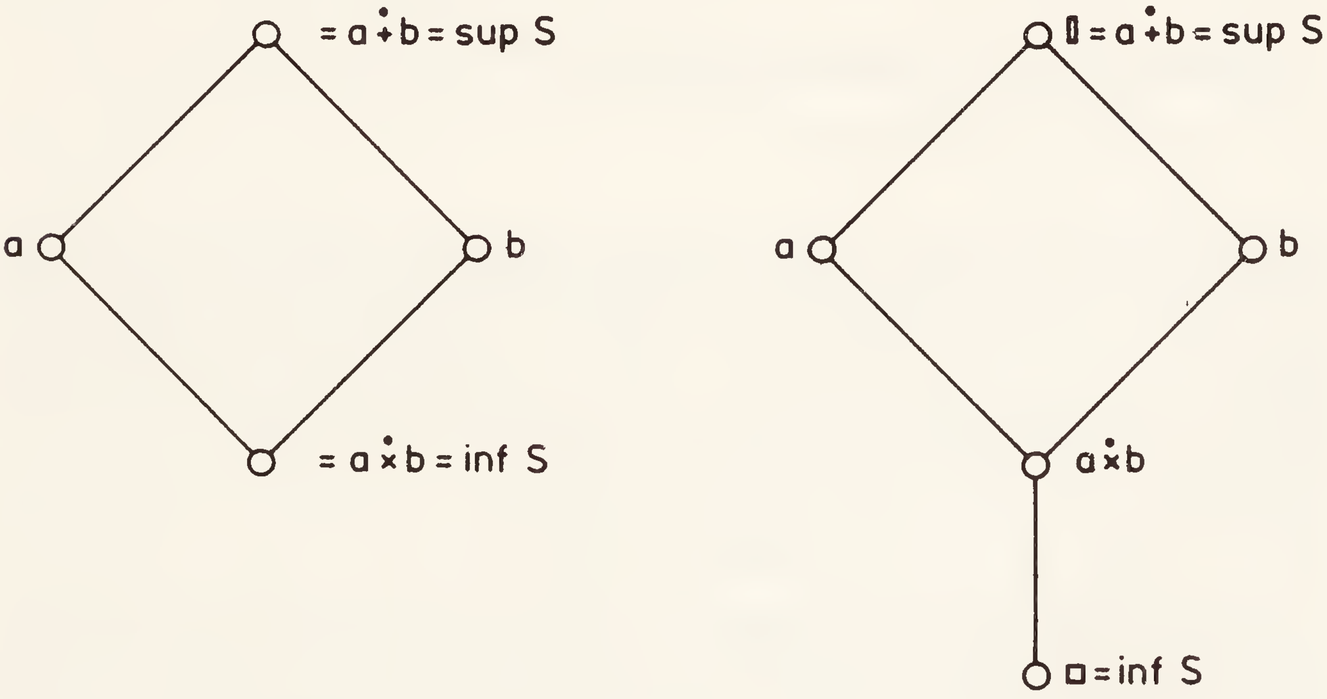
Our association theory bears a superficial resemblance to the ontology proposed by Lesniewski in 1916 (see Sobociński, 1954–55) and worked out in different directions by Tarski (1927), Leonard and Goodman (1940), Lejewski (1967), and a few others. The main differences are these: (a) our theory is much simpler than the others, if only because it adopts a standard mathematical formalism (namely semigroup theory) and does not eschew the language of set theory – which Lesniewski was careful to avoid; (b) whereas the Lesniewski theory (called *mereology*) and its offspring form the totality of contemporary nominalist ontology, our association theory constitutes only the foundation stone of a vast edifice containing items such as properties and kinds, which are anathema to the nominalist; (*c*) our association theory, and the assembly theory to follow, have been conceived not only as philosophical theories but also as theories utilizable in the foundations of science, as we shall explain in Sec. 4.

2. ASSEMBLY

2.1. *Juxtaposition and Superposition: Intuitive Idea*

The notion of association, elucidated in the previous section, is extremely general: there are many ways in which two substantial individuals can associate. We shall presently introduce two specific concepts of association: those of juxtaposition or physical sum, and of superposition or physical product. Two things placed side by side add or juxtapose, while two fluids, when mixed, superpose. We shall formalize these modes of association as so many binary operations in the set S of substantial individuals, designating juxtaposition by '$\dot{+}$' and superposition by '$\dot{\times}$'.

Example 1. Let the universe be cut into two mutually detached parts named a and b, i.e. ▯ $= a \dot{+} b$, with $a \dot{\times} b = \square$. The set $S = \{a, b, \square,$ ▯$\}$ is a complete lattice, since all the members of S can join pairwise in two ways without enlarging S: look at Figure 1.3(*a*).



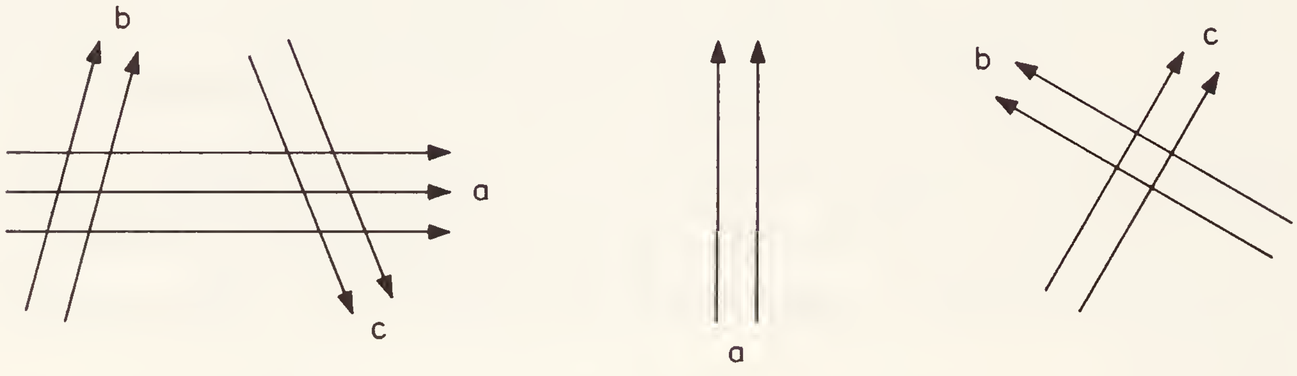
(a) (b)

Fig. 1.3. Two examples of juxtaposition and superposition.

Example 2 Let the universe consist of two superposed things, i.e. ▯ $= a \dot{+} b$, with $a \dot{\times} b \neq \square$. The set $S = \{a, b, a \dot{\times} b, \square,$ ▯$\}$, which includes the previous set, has the lattice structure too: look at Figure 1.3(*b*).

Example 3 Consider three individuals a, b, c, such that a superposes with both b and c. Suppose furthermore that the individuals are fields: look at Figure 1.4(*a*). It is seen that superposition distributes over juxtaposition.

Example 4 Juxtapose a field a and a field $b \dot{\times} c$. The result is that juxtaposition distributes over superposition. See Figure 1.4(*b*).



(a) (b)

Fig. 1.4. (*a*) The total entity is $a \dot{\times} (b \dot{+} c) = (a \dot{\times} b) \dot{+} (a \dot{\times} c)$. (*b*) The total entity is $a \dot{+} (b \dot{\times} c) = (a \dot{+} b) \dot{\times} (a \dot{+} c)$.

So much for the intuitive background.

2.2. *Formalization*

The examples discussed in the previous subsection suggest that we rope in the algebraic concept of a complemented distributive lattice. (Cf. Rutherford, 1965.) A lattice is a set S equipped with two binary associative operations, $\vee$ and $\wedge$, that are commutative and tied together by the so-called absorption laws

$$x \wedge (x \vee y) = x \vee (x \wedge y) = x.$$

A uniquely complemented lattice is a lattice with universal bounds $\square$ and ▯ and such that every element x of S has a single complement x' such that $x \vee x' =$ ▯ and $x \wedge x' = \square$. And a distributive lattice is one in which every operation distributes over the other. Finally a Boolean lattice is a lattice of idempotents which is both complemented and distributive.

More explicitly: the structure $\mathscr{L}=\langle S, \vee, \wedge, ', \square, ▯\rangle$ is a *Boolean lattice* iff, for all x, y and z in S, the following equalities hold:
Associativity $x \vee (y \vee z)=(x \vee y) \vee z, \quad x \wedge (y \wedge z)=(x \wedge y) \wedge z$
Commutativity $x \vee y=y \vee x, \quad x \wedge y=y \wedge x$
Absorption $x \vee (x \wedge y)=x, \quad x \wedge (x \vee y)=x$
Distributivity $x \vee (y \wedge z)=(x \vee y) \wedge (x \vee z), x \wedge (y \vee z)= =(x \wedge y) \vee (x \wedge z)$
Least element $x \vee \square=x, x \wedge \square=\square$
Last element $x \vee ▯=▯, x \wedge ▯=x$
Inverse $x \vee x'=▯, x \wedge x'=\square$.

Lattice theory together with the ontological interpretation we sketched in the previous subsection constitutes the ontological theory to be called *assembly theory*. The foundation of this theory is

POSTULATE 1.5 The structure $\mathscr{L}=\langle S, \dot{+}, \dot{\times}, ', \square, ▯\rangle$, where S is a non-empty set, $\square$ and $▯$ selected elements of S, $\dot{+}$ and $\dot{\times}$ binary operations in S, and $'$ a unary operation in S, is a uniquely complemented and distributive lattice of idempotents obeying the following additional conditions:

(i) S is the set of all substantial individuals;

(ii) $\square$ is the null individual and $▯$ represents the world;

(iii) for any individuals x and y, $x \dot{+} y$ represents the juxtaposition (physical addition) of x and y, whereas $x \dot{\times} y$ represents the superposition (physical product) of x and y;

(iv) the inverse (complement) of an individual x, i.e. the individual x' such that $x' \dot{+} x=▯$ and $x \dot{\times} x'=\square$, represents the environment or external world of x.

Remark 1 The unary operation (inversion or complementation) is defined for every individual entity. The world and the null individual are the inverses of one another. (The world, then, has no environment.) Needless to say, x' is not the opposite of x but just its complement to the world. *Remark 2* The two binary operations are defined for every pair of individuals. But whereas the juxtaposition of any two non-null entities is a third non-null individual, their superposition may be null – in which case they are detached, as we shall see in Definition 1.8. *Remark 3* At first sight superposition can result in annihilation, and wave interference would seem to illustrate this process. This is not so: two fields can superpose destructively over part of the region where they exist but not over the whole of it – otherwise there would be no conservation of energy and angular momentum. *Remark 4* When two or more individuals associate additively to form a whole they retain their identities. Consequently an aggregate, as treated by the present theory, is fully characterized by its composition. *Remark 5* By distinguishing two modes of association, i.e. juxtaposition and superposition, assembly theory allows us to perform a finer analysis of wholes than association theory did. Even so, $\dot{+}$ and $\dot{\times}$ are insufficient for scientific purposes, as things are endowed with a number of properties that determine their mode of coupling, which can be extremely complex. Just think of chemical and social bonds.

The preceding axiom will allow us to refine a few interesting notions.

2.3. *Definitions*

Since we now distinguish two modes of association we may expect roughly twice as many derived concepts as in association theory. Here are some.

DEFINITION 1.8 Let x and y be two distinct individuals. Then

(i) x and y are *detached* (or *separate*) iff their superposition is nought:

$$x \mathrel{\rceil\!\lfloor} y =_{df} x \dot{\times} y = \Box;$$

(ii) x and y *intersect* iff they are not detached:

$$\underset{\smile}{\overset{\frown}{x\ \ y}} =_{df} x \dot{\times} y \neq \Box.$$

Remark No spatiotemporal notions are involved in these definitions. On the contrary, the above notion of separateness will occur in the definitions of spatial concepts in Ch. 6.

DEFINITION 1.9 An individual $z \in S$, other than $\Box$,

(i) is *composed additively* of the individuals x and y iff $z = x \dot{+} y$;

(ii) is *composed multiplicatively* of the individuals x and y iff $z = x \dot{\times} y$.

Examples The contents of a room is composed additively of all the pieces of furniture in it. The gravitational field in a room is composed multiplicatively of the terrestrial field and the fields attached to the house.

DEFINITION 1.10 A substantial individual is *composite* (or *complex*) iff it is composed additively of individuals other than itself and the null individual. Otherwise it is *simple* (or atomic or basic).

DEFINITION 1.11 If x and y are substantial individuals, then x is a *part* of y iff x adds nothing to y:

$$x \sqsubset y =_{df} x \dot{+} y = y.$$

Justification The part-whole relation $\sqsubset$ has the desired properties: reflexivity, asymmetry, and transitivity. It is then a partial order relation. If a and b are additive components of c, they are also parts of c, i.e. they precede c in the order generated by $\sqsubset$. On the other hand if a and b are multiplicative components of c, they follow the individual $c = a \dot{\times} b$.

As in Sec. 1.2, we define the composition of a substantial individual as the collection of its parts: we just copy Definition 1.6. The complexity of an individual may be defined as the numerosity of its composition, i.e.

$$\text{For any } x \in S, \text{ the } \textit{complexity} \text{ of } x =_{df} |\mathscr{C}(x)|,$$

where $|S|$ stands for the cardinality of S. However, this is a coarse measure of ontic (nonconceptual) complexity, as it accounts only for the number of components or parts not for the manner of their coupling. But since we do not yet have the concept of coupling this is all we can do for the moment.

Concerning the complexity of the world we are faced with a choice among the following hypotheses:

H1 *Finitism*: The complexity of $\Box$ is finite.
H2 *Denumerable infinitism*: The complexity of $\Box$ = aleph zero.
H3 *Infinitism*: The complexity of $\Box$ = the power of the continuum.

However, we need not make any such choice at this point.

We now have two different concepts of aggregation:

DEFINITION 1.12 Let $T \subseteq S$ be a set of substantial individuals. Then

(i) the *additive assemblage* of T, or $[T]$ for short, is the supremum (l.u.b.) of T: $[T] = sup\, T$;

(ii) the *multiplicative assemblage* of T, or (T) for short, is the infimum (g.l.b.) of T: $(T) = inf\, T$.

Remark 1 Actually we have just introduced two different assemblage functions :[ ], ( ): $2^S \to S$, each of which takes a set of entities into an individual entity. By assuming Postulate 1.4 we need only the restrictions of these functions to the power set of the set of basics.

Remark 2 The two bounds, *sup T* and *inf T*, are of course defined with respect to the part-whole relation. They are such that, for all $x \in T$,

$x \sqsubset [T]$ and $(T) \sqsubset x$. Besides, if $x \sqsubset y$ for all $x \in T$, then $[T] \sqsubset y$. And, if $y \sqsubset x$ for all $x \in T$, then $y \sqsubset (T)$. *Remark 3* If $T \subset S$ is a finite set of substantial individuals then its additive aggregation is the juxtaposition of its members, whereas its multiplicative assemblage is their superposition. I.e., if $T = \{x_1, x_2, \ldots, x_n\}$, then

$$[T] = x_1 \dotplus x_2 \dotplus \ldots \dotplus x_n, \quad \text{and} \quad (T) = x_1 \dot{\times} x_2 \dot{\times} \ldots \dot{\times} x_n.$$

Remark 4 If T happens to contain a pair of detached entities, then its multiplicative aggregation is null, i.e. $(T) = \Box$. This holds, in particular, for any set of individuals containing the complement of one of them.

Finally we introduce a notion of decomposability. The intuitive idea is this. Let z be a substantial individual composed, no matter how, of two or more individuals. The composite individual will be said to be decomposable into x and y if these are different individuals such that $z = x \dotplus y$ and $x \rceil y$. The components x and y need not be those that make up the original whole z before disintegrating into the separate entities x and y. Nor need they coexist with z: they may emerge in the act of decomposition. The exact and general notion of decomposability is this:

DEFINITION 1.13 Let x be a substantial individual. Then x is *decomposable* iff there is a set $T \subset S$ such that $[T] = x$ and $(T) = \Box$.

Remark From an operationist point of view only effective decomposition counts as an indicator of structure: if no way is known of actually splitting an entity then the latter is declared simple. Counterexample: current particle theory assigns the neutron a structure, hence a composition, but so far the neutron has withstood all efforts at decomposing it. Only the converse of the operationist thesis holds, namely: Anything effectively decomposable has a structure.

DEFINITION 1.14 A substantial individual is *uniquely decomposable* iff it is decomposable in but one way – i.e. if there is a single set $T \subset S$ such that $[T]$ equals the individual concerned and $(T) = \Box$.

Remark Here again unique decomposability is not to be mistaken for the single known way of effectively decomposing a thing. For example, so far the deuteron has been decomposed only into a neutron and a proton; but this does not exclude the possibility of decomposing it into a different set of products.

2.4. *Some Consequences*

We can retrieve a number of results obtained in association theory, since $\dotplus$ and $\circ$ are practically the same. In addition we can obtain a number of new theorems. But we shall exhibit only a few. Firstly

THEOREM 1.7 Every substantial individual is the additive aggregation of its parts. I.e., for every $x \in S$,

$$x = [\mathscr{C}(x)] = sup\{y \in S \mid y \sqsubset x\}.$$

COROLLARY 1.8 Every substantial individual with a finite composition is the juxtaposition of its parts. I.e., for any $x \in S$,

$$\text{If } \mathscr{C}(x) = T \equiv \{x_1, x_2, \ldots, x_n\} \quad \text{then}$$

$$x = [T] = x_1 \dotplus x_2 \dotplus \ldots \dotplus x_n.$$

COROLLARY 1.9 The world is the additive assemblage of all substantial individuals: $\Box = [S] = sup\, S$.

THEOREM 1.8 Every entity is part of some substantial individual: If $x \in S$ then there is at least one $y \in S$ such that $x \sqsubset y$.

Proof Take an arbitrary pair $a, b \in S$ and form the third individual $c = a \dotplus b$. This individual exists and is unique since, by hypothesis (Postulate 1.5), S forms a lattice. And, by definition of the part-whole relation, $a, b \sqsubset a \dotplus b$.

COROLLARY 1.10 Every entity is composite either properly or improperly. I.e., for every $x \in S$, $(\exists y)(y \in S \;\&\; y \sqsubset x)$.

Proof Follows from Theorem 1.8.

Remark 1 That an individual is improperly (or trivially) composite means of course that it consists just of itself. The assumption that every individual we might stumble upon might prove to be properly composite, is a powerful heuristic assumption that must not be confused with the preceding corollary. *Remark 2* Nor should the previous result be mistaken for the statement that S has a least element other than the null element and that would be part of every other individual.

We shall now use the notions of decomposability (Definition 1.13) and unique decomposability (Definition 1.14) to obtain a couple of interesting results.

THEOREM 1.9 All substantial individuals are decomposable. Decomposition includes the trivial decomposition of an individual into itself and the null entity $\square$.

Proof Let y be a part of $x \in S$ and take the environment y' of y: $y \dotplus y' = \square$ and $y' \dot{\times} y = \square$. Call $z = y' \dot{\times} x$ the superposition of y' with x. This is what completes y to form x: $y \dotplus z = x$, with $y \dot{\times} z = \square$. We have effected the decomposition of an arbitrary entity x into y and z.

THEOREM 1.10 No composite individual is uniquely decomposable. Equivalently: For all $x \in S$ other than $\square$, x is uniquely decomposable iff x is simple (i.e. decomposable only trivially: $x = x \dotplus \square$).

Proof (i) The *only if* part: let $y \sqsubset x$. Then $x = y \dotplus (y' \dot{\times} x)$ is a decomposition. In order to be unique, this must be $x = x \dotplus \square$ or $x = \square \dotplus x$. Thus either $y = x$ or $y = \square$. (ii) The *if* part: if $x = y \dotplus z$ and x is simple, then $y = \square$ or $y = x$ and also $z = \square$ or $z = x$. There are two possibilities: if $y = \square$ then $z = x$ and we have the decomposition $x = \square \dotplus x$. And if $y = x$ and also $z = x$ then $y \dot{\times} z = x \neq \square$. But this is a contradiction, so we must have $z = \square$ and thus the decomposition $x = x \dotplus \square$.

Now the conservation law or Epicurus'

THEOREM 1.11 If $x, y \in S$ are not the null individual, then

(i) their juxtaposition does not end up in nought: $x \dotplus y \neq \square$;

(ii) provided they are not detached their superposition does not ensue in nought either: $x \dot{\times} y \neq \square$;

(iii) no being comes out of nonbeing either by juxtaposition or by superposition: $[\square] = \square$ and $(\square) = \square$.

Finally a couple of theorems concerning separation or detachment (Definition 1.8).

THEOREM 1.12 Separation is hereditary, i.e. the parts of detached entities are detached. That is, for any substantial individuals x and y,

$$x \lfloor y \,\&\, u \sqsubset x \,\&\, v \sqsubset y \Rightarrow u \lfloor v.$$

Proof By definition of detachment and of part,

$$x \dot{\times} y = (x \dotplus u) \dot{\times} (y \dotplus v) = \square.$$

Performing the indicated operations and recalling the distributivity of $\dotplus$ over $\dot{\times}$,

$$x \dot{\times} y = (x \dot{\times} y) \dotplus (x \dot{\times} v) \dotplus (u \dot{\times} y) \dotplus (u \dot{\times} v) = \square.$$

By Theorem 1.11 every addend must be null. In particular, $u \dot{\times} v = \Box$, which amounts to $u \lfloor v$.

Remark This theorem shows that our detachment relation $\lfloor$ for entities is analogous to the separation relation for sets (Wallace, 1941).

THEOREM 1.13 Detachment is additive. I.e., if two substantial individuals are detached from a third so is their juxtaposition: for any three entities $x, y, z \in S$,

$$x \lfloor z \;\&\; y \lfloor z \Rightarrow (x \dotplus y) \lfloor z.$$

Proof Suppose the consequent is false. This assumption is equivalent to $(x \dotplus y) \dot{\times} z \neq \Box$, which expands into $(x \dot{\times} z) \dotplus (y \dot{\times} z) \neq \Box$. In turn the last statement amounts to $x \dot{\times} z \neq \Box \vee y \dot{\times} z \neq \Box$. But the first disjunct amounts to $\neg(x \lfloor z)$ and the second to $\neg(y \lfloor z)$, contrary to hypothesis. So, the consequent is true.

Remark In the theory of the Wallace separation spaces (Wallace 1941) additivity is postulated not demonstrated.

2.5. *Atoms and Levels*

The general notion of composition (Definition 1.6) has got to be supplemented with another, more specific notion, which will prove to be more useful than the former, namely that of composition relative to a certain set or level of entities. Thus in the case of an animal society regarded as a whole, we are interested in the set of its components not in the full set of its parts, such as the cells of the animals, even less the atomic components of their cells. That is, we want to know what the "relative" atoms of the whole are. The idea we want is this:

DEFINITION 1.15 Let $A \subset S$ be a set of entities. Then the *A-composition* (or *composition at the A level*) of a substantial individual $x \in S$ is the set of parts of x belonging to A:

$$\mathscr{C}_A(x) = \mathscr{C}(x) \cap A = \{y \in A \mid y \sqsubset x\}.$$

Example The molecular composition of a body of water is the set of its H_2O molecules. On the other hand the atomic composition of the same body is the set of H and O atoms that compose it.

We shall make ample use of this notion in Vol. 4, Ch. 7 on systems, where the composition of a system will be taken to be a certain

A-composition. We shall want the composition of a system to equal the union of the compositions of its subsystems. However, in general the composition of a complex individual is not equal to the union of the compositions of its components: see Figure 1.5. On the other hand the

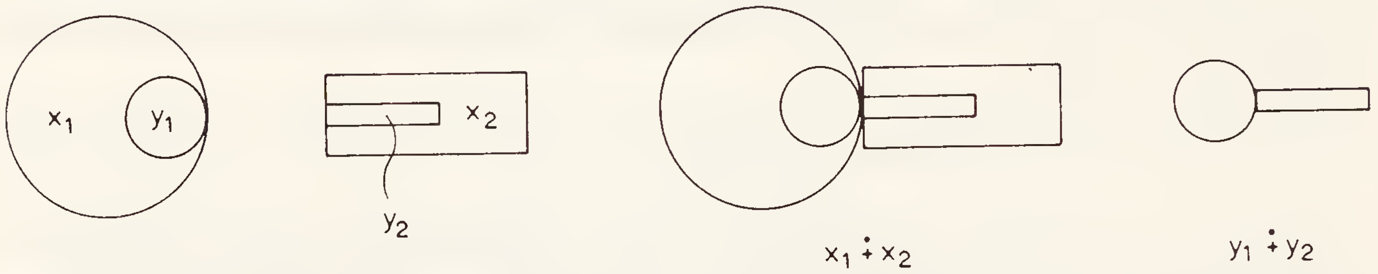


Fig. 1.5. The composite individual $x_1 \dotplus x_2$ has a part, namely $y_1 \dotplus y_2$, which is neither a part of x_1 nor a part of x_2. In general $\mathscr{C}(x_1) \cup \mathscr{C}(x_2)$ is a proper part of $\mathscr{C}(x_1 \dotplus x_2)$. Not so with the A-composition of a compound individual: see Theorem 1.14.

A-composition does have the desired property and this is one reason for preferring it over the general notion of composition. Indeed, we have

THEOREM 1.14 Let $A \subset S$ be a set of substantial individuals (level A atoms), and $x_1, x_2 \in S$ two substantial individuals. Then

$$\mathscr{C}_A(x_1 \dotplus x_2) = \mathscr{C}_A(x_1) \cup \mathscr{C}_A(x_2).$$

Proof If $x \in \mathscr{C}_A(x_1)$ then x is an A-atom (a member of A) and therefore $x \sqsubset x_1 \sqsubset x_1 \dotplus x_2$. Therefore $x \in \mathscr{C}_A(x_1 \dotplus x_2)$. Similarly $x \in \mathscr{C}_A(x_2) \Rightarrow x \in \mathscr{C}_A(x_1 \dotplus x_2)$. Thus $\mathscr{C}_A(x_1) \cup \mathscr{C}_A(x_2) \subseteq \mathscr{C}_A(x_1 \dotplus x_2)$. Now the converse implication. If $x \in \mathscr{C}_A(x_1 \dotplus x_2)$ then $x \sqsubset x_1 \dotplus x_2$, which is equivalent to $x \dot{\times} (x_1 \dotplus x_2) = x$. By distributivity it follows that $x \dot{\times} x_1 \dotplus x \dot{\times} x_2 = x$. But x is an A-atom and therefore we must have $x \dot{\times} x_1 = x$ or $x \dot{\times} x_2 = x$, i.e. $x \sqsubset x_1$ or $x \sqsubset x_2$. We have just shown that $\mathscr{C}_A(x_1 \dotplus x_2) \subseteq \mathscr{C}_A(x_1) \cup \mathscr{C}_A(x_2)$. This, conjoined with the previous result, yields the desired proof.

The entities of kind A, or A-atoms, may be simple (i.e. members of the set B of basics) or complex. To qualify as a member of a given set A, an entity must be able to associate with similar entities to form complex entities of a different kind. But we do not have yet the notion of a natural kind, which will be introduced in Ch. 4, Sec. 3.3. Therefore for the time being we shall have to make do with a presystematic or intuitive notion of an A-atom. The same holds *a fortiori* for the notion of a level. All we can say at this time is that we may split the totality S of entities into a certain number n of disjoint sets A_i, each member of which is composed – additively or multiplicatively – by entities of the next lower level A_{i-1}. The levels hypothesis may then be formulated as follows:

$$S = \bigcup_{i=1}^{n} A_i, \quad \bigcap_{i=1}^{n} A_i = \varnothing, \text{ with } A_i = \{a_i^n \mid n \in \mathbb{N}\}$$

and

$$a_i^n = [\alpha_{i-1}^n] \quad \text{or} \quad (\alpha_{i-1}^n), \quad \text{where } \alpha_{i-1}^n \subseteq A_{i-1}.$$

The matter of (relative) atoms and levels of organization will be taken up again in Vol. 4, Ch. 8.

2.6. *Alternative Formalizations*

Alternative formalizations of the notions of juxtaposition and superposition have been explored. One of them is in terms of ring theory, i.e. it consists in assuming that $\langle S, \dot{+}, \dot{\times}, \square, \text{▯}\rangle$ constitutes a ring of idempotents (Bunge, 1971a, 1973a). In this construal one must abstain from interpreting the unit element of the ring as the world, for in this case the unit is not the last element. This is not a fatal flaw: the defect consists in leaving one primitive devoid of ontological interpretation. What may constitute a defect is that in ring theory $\dot{\times}$ distributes over $\dot{+}$ but not conversely. As a compensation the absorption laws, which seem idle in the ontological interpretation of lattice theory, do not occur in ring theory.

A second alternative formalization consists in a certain interpretation of Boolean algebra (Bunge, 1967b). The two formalizations, though different, are equivalent. In fact by a famous theorem of Stone's, a commutative Boolean ring $\langle S, +, \cdot, 0, 1\rangle$ determines a Boolean algebra $\langle S, \cup, \cap, 0, 1\rangle$ with meets and joins defined as follows:

$$x \cup y = x + y + x \cdot y, \qquad x \cap y = x \cdot y.$$

We may adopt either of these alternative formalizations in the case where the reference class is a proper subset of the total set S of substantial individuals. In this case the unit will not be interpreted as the world. The ring theoretic construal may be preferred when this condition is fulfilled and superposition is not required to be commutative, as is often the case with two parts of a machine.

Needless to say a third alternative, having none of the defects of the preceding ones, should be welcome. The same applies to every one of our ontological theories: different formalizations of the same intuitions, as well as alternative intuitions, should be explored.

2.7. *Concluding Remarks*

Our two theories of substance – namely association theory and assembly theory – systematize the intuitive ideas of a physical object and of the association of concrete individuals. We have not called any of our theories *mereology* – as we should if we had followed in Lesniewski's path – because they achieve more than an elucidation of the part-whole relation. If we were interested just in the latter concept we could adopt a simpler theory, namely that of partially ordered sets. Indeed, assuming that the structure $\langle S, \sqsubset \rangle$ is a poset and adding the obvious interpretations of S and $\sqsubset$, we would get an exact characterization of the notion of a part. And the latter would in turn allow us to define a few other concepts, notably those of detachment and composition. However, such a theory would be too poor: for one thing it would not clarify any notion of association – and this is what our theories of substance are all about.

Nor have we called our theories *calculi of individuals* – as Leonard and Goodman (1940) called their version of Lesniewski's mereology. One reason is that every first order mathematical theory is a calculus of individuals: whether it is a calculus of physical individuals is another matter. Another reason is that, unlike Lesniewski and his fellow nominalists, we draw a radical distinction between physical and conceptual individuals and reject the thesis that a single theory should be able to account for both kinds of individual. In particular our part-whole relation applies only to ontic individuals, i.e. $\sqsubset$ is defined on $S \times S$. More on this matter in the next section.

3. ENTITIES AND SETS

3.1. *The Null Individual and the World*

Both association theory and assembly theory involve the greatest fiction (□) as well as the greatest reality (▯). The null individual has been considered by Carnap (1947), Martin (1965) and a few others. It has usually been characterized by analogy with the empty set, namely as "that thing which is a part of every thing". This characterization is not quite correct. Firstly $\square$ is not a thing but a concept. Secondly the characterization is deficient: the null body is also the null non-body. (Besides, according to the aether theories, the aether, though not null, was supposed to be part of every thing.)

The null individual, or nonbeing, is best characterized in terms of either association or assembly, i.e. either as (i) that which, associated with any substantial individual, yields the latter, or as (ii) that which, juxtaposed to any thing leaves it unchanged and, superposed to it, annuls it. In either case $\square$ is in S and is part of every member of S. (The claim that $\square$ belongs to every set, even the empty set, leads to contradiction.) All these are definite properties of $\square$. That is, according to the theories formulated in the previous sections, it is false to say that the null individual has no properties – as false as to say that it has substantial (physical, nonconceptual) properties. Which reminds one of Lucy's verdict: "For a nothing, Charlie Brown, you're really something" (Schulz, 1967).

What is true is that the null individual simply does not exist physically: it is a fiction introduced to get a smooth theory. We need this pretence not only in ontology but also in the foundations of science – which is not surprising because the two intersect. Thus the vacuum, i.e. the null field, is assigned a number of properties in field theories – such as a refractive power equal to unity. Moreover it is convenient to introduce several null individuals, one for every natural class. For example, in optics we need the void, or null body b_0, and darkness, or the null light field l_0. The reference class of a statement about a light beam l in vacuum is $\{l, b_0\}$, while a statement about a body b in darkness is about $\{l_0, b\}$, and one about the dark void concerns $\{b_0, l_0\}$ (Bunge, 1966).

We need the null entity to think about real entities but we cannot use it to build the latter. Indeed, unlike the empty set in von Neumann's reconstruction of number theory, the null individual is not the point of departure of our cosmology but only a very modest component of it, which discharges a theoretical function but has no counterpart in reality. In other words, entities are not constructible out of the nonentity. This is in sharp contrast with the central role Hegel (1812, 1816), Peirce (1892–93, 6.217) and Heidegger (1927) have assigned nonbeing. (Incidentally none of them seems to have distinguished the nonentity, or nonbeing, i.e. $\square$, from nothingness or the empty set $\varnothing$. Yet $\square$ and $\varnothing$ are radically different, i.e. they have different properties even though none of them is substantial or real.) Far from taking nonbeing as our point of departure, our theories of substance involve the theses of Epicurus (*apud* DeWitt 1955) and Lucretius (55 B.C.?) (*a*) that no thing emerges *ex nihilo* and (*b*) that no thing goes over into nought (Theorem 1.11). Our point of departure is the real world as known to the science of the day: wishing to understand the world we dismantle it in thought and oversimplify its ingredients as well as the manner in which they are actually assembled. We thus obtain a concept of substance, or matter, which, when enriched with properties, will ensue in the concept of a concrete thing (Ch. 3).

3.2. *Entities and Concepts*

While the null individual, a concept, *is* (identical with) □, the world (or reality) is said to be *denoted by* the symbol '▯'. In our theories of substance the world is an individual but not just one more individual: it is the entity that contains as components all other entities. But this is as far as our theories of substance go: they give no details about the structure of ▯. On the other hand they do assign definite structures (semilattice in one case, lattice in the other) to the furniture of the world, i.e. S.

We emphasize that S, the set of all entities, is not the same as the aggregate or whole $[S]$ composed of all physical objects. The difference between the concept S and the entity denoted by $[S] = ▯$ illustrates the construct/thing dichotomy. To us the basic dichotomy in any set of objects is not that between individuals and sets but that between physical objects and conceptual ones. That is, we assume that every class O of objects is split into a class C of constructs and a class $T \subseteq S$ of substantial individuals:

$$O = C \cup T, \quad \text{with} \quad C \cap T = \varnothing.$$

Sets, relations and functions, as well as algebraic and topological structures, rank as constructs in our philosophy and, as such, they do not satisfy any theory of substance.

Whether an object is an individual or a set is of interest to ontology: if the former then it may be either physical or conceptual, if a set it can only be a concept. But whether a *conceptual* object (a construct) is an individual or a set is metaphysically and even mathematically relatively unimportant. This is for the following reasons. First, what is a (conceptual) individual in one theoretical context may become a set in another and conversely: it all depends on the fineness of our analysis. Second, some set theories draw hardly any distinction between sets and individuals. Nevertheless there is one assumption (von Neumann's D axiom) that does draw the line between the two categories by stipulating that every chain of the form ... $x \in y \in z$ has a stop. (For every nonempty set there is a member none of whose members is a member of the given set. Such a member may be regarded as an individual with respect to the given set. For example, in $A = \{\{a\}\}$ the element a is such an individual, the ascending chain being $a \in \{a\} \in A$.) Even if this axiom is not accepted, the individual/set distinction is kept in everyday mathematics. We shall keep it.

We are now in a position to clarify the difference between our ontology and Platonism with regard to individuals. (The difference with respect to properties will be discussed in Ch. 2, Sec. 5.) Suppose we agree to consider, or rather toy with, a universe composed of a finite number of atoms, or simple individuals. We start by inquiring whether these individuals are conceptual or concrete. (On the other hand Goodman, 1956, p. 16 regards this distinction to be "vague and capricious".) If they are conceptual then we let the Platonist take over, for that is his field, though where he believes in the autonomous existence of such ideas we just pretend that to be so for the sake of expediency. He should feel free to build not just the power set of the given set but also any powers of that power set if he needs them. On the other hand if the individuals are concrete then we give the Platonist a push and do our own reckoning. The upshot is this: in addition to the original atoms we get all their partial (binary, ternary, etc.) associations. And on top of these we shall have all the novel things that may ensue from such combinations, since we know from high energy physics that a couple of elementary particles may collide to produce a large number of new individuals. Thus we come up with a universe that, in numerosity, is infinitely smaller than the Platonic universe though appreciably larger than the universe of unchangeable and unreactive atoms envisaged by mechanism.

3.3. *Existence and Individuation*

Russell taught us that the existential quantifier ∃ cannot occur by itself but must be applied to "something described" such as Ax, where A is a unary predicate. Hence at first sight it would seem that our substantial individuals do not exist, for they have no peculiarities. However, they do share the definite property of associating with one another. That is, every member of $\langle S, \circ, \square \rangle$ is characterized by the predicate A defined by

$$Ax =_{df} (\exists y)(\exists z)(y \in S \;\&\; z \in S \;\&\; x = y \circ z \vee y = x \circ z \vee z = x \circ y).$$

Hence we may prefix (x) to this propositional schema, obtaining the proposition "All individuals associate".

Moreover we can individuate those entities with a known composition. That is, we can form definite descriptions of the form

"The member x of S with composition $\mathscr{C}(x)$".

In traditional terms: our *principium individuationis*, at this early stage, is composition. Obviously this principle fails us for simples since, with the exception of $\square$, they are devoid of peculiarities other than that they can associate. (Therefore we cannot use Quine's method of treating all singular terms as definite descriptions: Quine (1950, Sec. 37). In order to use this method we must have stock of predicates large enough to uniquely characterize every individual, and this will have to wait till next chapter.) Besides, the above principle does not even allow us to distinguish between two molecules with the same composition. But this is only a temporary difficulty, as a much finer *principium individuationis* will be adopted in Ch. 2, Sec. 3.2.

Finally a word or two about substantial existence as distinct from existence in general. The existence of simple individuals can be ascertained only by scanning the entire set S. If a belongs to S and is not the null individual we say that a exists. That is, we explicate "a exists" as "$a \neq \square \;\&\; a \in S$". This concept of existence can be called *bare substantial existence.* It suffices in the theory of substance, where the only individuating property is composition. (More on physical existence in Ch. 3, Sec. 4.4.) Note that this is a notion of ontic or physical existence to be distinguished from that of conceptual existence. In formal science to say that an individual a exists it suffices to assume or to prove that a belongs to some set that has been satisfactorily characterized. In ontology we have no use for arbitrary sets except perhaps as auxiliaries devoid of ontological import: here to exist (substantially, physically) is to have a number of substantial properties, among them that of associating with other nonconceptual objects. So, *pace* Wittgenstein (1969, Sec. 35), it makes perfectly good sense to state that there are substantial (physical) objects. Moreover ontology happens to be concerned precisely with such objects. (Recall Introduction, Sec. 2.)

4. CONCLUDING REMARKS

One of the necessary conditions for an ontological theory to qualify as scientific is its relevance to science. (Cf. Introduction.) Association theory and assembly theory comply with this condition: they are relevant to any scientific or technological theory concerned with complex systems such as atoms, cells, machines, brains, or organizations. In particular any such theory employs a concept of association and therefore should welcome an elucidation of the latter. However, in the vast majority of scientific writings only an intuitive concept of association is employed. It is only when the foundations of a scientific theory are carefully laid down that $\circ$, $\dot{+}$ or $\dot{\times}$ are apt to occur explicitly, because they are needed. For example, the very statement of the hypothesis that the electric charge Q is additive involves $\dot{+}$:

If x and y are bodies, and $x \neq y$,
then $Q(x \dot{+} y) = Q(x) + Q(y)$.

Similarly with the intensity E of the superposition of two electric fields:

If x and y are fields, and $x \neq y$,
then $E(x \dot{\times} y) = E(x) + E(y)$.

If only for this reason, i.e. because it is needed, it is mandatory to include some version of assembly theory among the ontological presuppositions of any scientific theory referring to complex systems (Bunge, 1967b, 1973b, 1974d).

Although the scientific method is analytic, from time to time the scientific community is swept by holistic winds – particularly when science fails to disclose the actual composition and inner workings of certain things. At such times the black box theories come in vogue and the mechanistic approach is decried or at least shelved. A characteristic of the black box approach is that it makes little or no use of the very concept of composition: it treats every component of a system, and occasionally even the whole system, as a structureless box. No doubt this strategy is wise as long as little is known, or need be known, about the components and their mode of association. But to advertise this *pis aller* strategy as optimal (as e.g. Chew, 1966, does) is suicidal. Indeed, it amounts to obliterating the difference between simple and composite, hence between microphysical and macrophysical, and it renounces the ideal of the most complete possible account of things. Worse: the strategy ignores whatever complexity has already been disclosed. Far worse: such a drastic reduction in our cognitive goals cannot be implemented in microphysics except by borrowing from translucid box theories such as field theory and ordinary quantum mechanics (Bunge, 1964; Regge, 1967). In short, the idea of composition, and the goal of finding out the precise composition and manner of association, are not "out": at most some grapes are still sour.

So much for the concept of substance or matter. Let us now tackle the problem of characterizing the concept of a property.

CHAPTER 2

FORM

The next step in our conceptual reconstruction of the notion of a real thing is to introduce the concept of a property, trait, or character. Every real thing has a number of properties: there is no formless substance except as a useful fiction. Nor are there pure forms hovering above matter. Every concrete or substantial property, such as moving, reacting, or remembering, is the property of some thing or other – bodies, reactants, brains, or what have you. Were this not so, science would not seek to find and interrelate properties of things, much less the patterns of the associations and changes of properties – i.e. the laws.

We cannot take the notion of a property for granted because it is far from clear. In any event no discipline other than ontology attempts to clarify the notion of a property as distinct from the various specific notions of property. Logic is certainly concerned with the notion of an attribute or predicate. But, as we shall see in Sec. 1, properties should be distinguished from the attributes representing them – except of course in the context of philosophical idealism, where they are conflated.

Nor can we accept uncritically the hypothesis that things have properties, particularly since it has been challenged by nominalist philosophers. In fact the latter wish to dispense with properties, which they regard as Platonic fictions, and attempt to reduce everything to things, their names, and collections of such (see e.g. Woodger (1951) and Cocchiarella (1972)). At the very least the nominalist will want to construe every property as a class of individuals, or of n-tuples of individuals – i.e. he will adopt an ontological interpretation of the semantic doctrine called *extensionalism*, which equates predicates with their extensions. But we have seen (Volume 1, Ch. 5, and Vol. 2, Ch. 10) that extensionalism is indefensible even in mathematics, if for no other reasons than that (*a*) some of the basic mathematical predicates, notably the membership relation, are not defined as sets, and (*b*) coextensive predicates need not be cointensive or have the same meaning. Therefore extensionalists can "deplore the notion of attribute" (Quine, 1963) without being able to dispense with it, much as the Victorians deplored sex while engaging in it.

If extensionalism fails so does its ontological interpretation, namely the reduction of the world to a collection of individuals devoid of properties. For one thing this ontology is incapable of explaining what makes a given individual different from any others: it has no *principium individuationis*. For another, nominalism commits the sin of reification, or the transformation of everything into things or collections of things. (Reification, a constant of archaic thinking, is occasionally found among scientists and philosophers. Thus the vulgar materialists of the 19th century held the world to consist of things of two sorts, *Stoff* or substance and *Kraft* or force. Later on the energeticists, wishing to dematerialize the world, proclaimed that everything is a manifestation of energy. And nowadays the followers of Whitehead maintain that there are no things but just events.) For those two reasons nominalism does not tie in with science, which regards things as individuals possessing properties, and handles the notion of a property as basic – so much so that every law statement is deemed to represent an invariant relation among properties of things.

We shall propose a construal of the notion of property that avoids the extremes of nominalism (or substantialism) and Platonism (or formism), and squares with the scientific conception.

1. PROPERTY AND ATTRIBUTE

1.1. *Difference between Property and Attribute*

All objects have properties. If the objects are conceptual or formal, their properties will be called *formal properties*, or *attributes* or *predicates* for short. If the objects are substantial individuals, their properties will be called *substantial properties*, or *properties* for short. Because every model of a substantial individual is built with concepts, it contains attributes or predicates; and insofar as the model represents a substantial individual, some of those attributes or predicates represent substantial properties.

In the case of a conceptual object, such as a set or a theory, the words 'attribute' and 'property' are exchangeable because a conceptual object has all the properties we consistently attribute to it. But in the case of a substantial individual we must distinguish a substantial property or objective trait from the corresponding attribute(s) if any. A substantial property is a feature that some substantial individual possesses even if we are ignorant of this fact. On the other hand an attribute or predicate is a feature we assign or attribute to some object: it is a concept. A predicate may conceptualize or represent a substantial property; but then it may not or it may do so poorly, i.e. with a large margin of error. On the other hand the possessing of a property is not a matter of truth or falsity; only our knowledge of properties can be more or less true or adequate. For this reason we distinguish

$\ulcorner$Substantial individual b *possesses* property $P\urcorner$.

from

$\ulcorner$Attribute A *holds* for $b\urcorner$,

or $\ulcorner A$ is *true* of $b\urcorner$, or $\ulcorner\mathscr{V}(Ab)=1\urcorner$,

where A is taken to represent P. Again: whereas the possessing (or the acquiring or the losing) of a property by a substantial individual is a fact often beyond our reach, our attribution of a property (via some predicate) is a cognitive act. In other words, whereas we control all predicates because we make them, we control only some properties.

Needless to say, no such difference between attributes and substantial properties is admitted by idealism: in such a philosophy the acquiring of a property coincides with its attribution. (Which is why the inadequacies of this philosophy are hardly realized in the realm of formal science.) Oddly enough naive realism, while recognizing the difference between attributes and substantial properties, asserts their one-to-one correspondence. Therefore the stand of those two poles of philosophy with respect to the theory of properties is rather close: whereas idealism asserts that the theory is the logic of predicates, naive realism is committed to the claim that the theory of substantial predicates is an ontological interpretation or model of the predicate calculus.

We proceed to approaching this matter from the vantage point of our semantics (Vols. 1 and 2).

1.2. *Attribute-Property Correspondence*

Surely we know properties only as attributes or predicates, i.e. as components of our views of things. Still, we must distinguish the object represented from its representations if we wish to account for discovery (or invention) and ignorance, truth and error. We do that by making statements of the form

$\ulcorner$Attribute A *represents* substantial property $P\urcorner$,

or $\ulcorner A \triangleq P\urcorner$ for short.

For example, in certain contexts angles represent inclinations, probabilities tendencies, densities crowdings, and so on. The attribute-property correspondence is a particular case of the knowledge-reality, or mind-world relation. This correspondence is not isomorphic because some attributes represent no substantial properties, others represent several properties, and finally some properties are represented by no attributes (because we ignore either or both) or by several predicates (often belonging to different theories of the same kind of thing).

The relation between properties and attributes may be construed as follows. Let $\mathbb{P}$ be the set of substantial properties and $\mathbb{A}$ that of attributes or predicates. Then the representation of properties by attributes is a function $\rho : \mathbb{P} \to 2^{\mathbb{A}}$ that assigns each property P in $\mathbb{P}$ a collection $\rho(P) \in 2^{\mathbb{A}}$ of attributes, so that the formula '$A \in \rho(P)$', for A in $\mathbb{A}$ and P in $\mathbb{P}$, is interpreted as *attribute A represents property P*, or $A \triangleq P$ for short. Details on the concepts of attribute (or predicate) and representation are found in Vol. 1, Ch. 3 and Vol. 2, Ch. 6.

The representation function is a correspondence between a proper subset of all the conceivable attributes (predicates) and the ill-defined set of all (known and unknown) substantial properties. That is, there are attributes with no ontic correlate. Among them we find membership in a set, the negative attributes, and the disjunctive ones. Let us discuss the three because they will show what a substantial property is not. First the matter of membership. In mathematics every member of an arbitrary (but well defined) set may be assigned the attribute of membership in that set. However, being an element of an arbitrary set does not count as a substantial property. For example, although I may decide to group my shoes together with President X's latest statement, such a common membership does not constitute an objective property shared by the two items. In short, not all sets are kinds.

Nor, *pace* a few distinguished philosophers (e.g. Bolzano (1851), Sec. 39 and Russell (1918), Secs. I and III), are "negative" properties possessed by factual items such as things or events. These have only "positive" properties (Bunge, 1974g). In the statement ⌜Neutrons are not electrically charged⌝ the negation affects the entire proposition ⌜Neutrons are electrically charged⌝: it denies the latter statement instead of attributing neutrons the negative property of not being charged, let alone that of being anticharged. We certainly need negation to understand reality and argue about it or anything else, but external reality wears only positive traits. In other words, even though a proposition of the form $\ulcorner \neg Ab \urcorner$ be true and indistinguishable from the proposition $\ulcorner (\neg A)b \urcorner$, there may not exist any real or substantial property represented by the negative attribute $\neg A$ and exemplified by the individual b. Consequently if attribute A represents substantial property P(i.e. $A \triangleq P$), then we can accept the truth condition (sufficient for qualitative attributes)

$\ulcorner \neg Ab \urcorner$ is true if and only if b fails to possess P.

but not

$\ulcorner \neg Ab \urcorner$ is true if and only if b possesses $\neg P$.

In short, negation is *de dicto* not *de re*: it is defined on the set $\mathbb{A}$ of attributes not on the set $\mathbb{P}$ of properties. And because there are no negative properties there are neither tautological nor contradictory substantial properties corresponding to tautological and contradictory attributes such as $A \vee \neg A$ and $A \;\&\; \neg A$ respectively.

Likewise there are no disjunctive attributes with ontic correlates. For example, there is no substantial individual possessing the property of being heavy *or* translucid, even though the proposition ⌜Lucite front doors are heavy or translucid⌝ is true. (In other words, we employ the disjunctive attribute "heavy or translucid" but do not have it represent a property possessed by any substantial individual.) On the other hand certain properties can conjoin. For example there are things both heavy and capable of synthesizing chlorophyll. (But of course other properties are mutually incompatible, i.e. cannot conjoin.) What the expression 'conjunction of properties' signifies will be seen in Sec. 3.4.

The preceding considerations, Byzantine though they may look, are necessary for building a theory of properties and they have far reaching philosophical consequences. A first consequence of the lack of isomorphism between properties and attributes is the ruin of naive realism, both in the form of the reflection theory of knowledge and of the picture theory of language. According to these views every factually true proposition depicts a fact. But if we admit just "positive" properties then we must rule out negative facts. And if we do not admit disjunctive properties then we must write off disjunctive facts, and *a fortiori* general facts. Of course we accept negative and disjunctive propositions as long as they are sufficiently true or at least promising, but that has nothing to do with accepting negative facts and disjunctive facts. Thus "There are no elephants in your pocket." is true precisely because this proposition fails to represent a situation in which there are elephants in your pocket. (On the other hand naive realism would have us interpret that statement as "Everything is either a nonelephant or unlocated-in-your-pocket".) Last, but not least, the attribute-property gap bars any attempt to read the predicate calculus in ontological terms, i.e. to get a theory of properties without tears. Therefore we must try and build a property calculus as distinct from the predicate calculus. To work.

2. ANALYSIS

2.1. *Property in General and Property of a Particular*

The expressions 'pure form', 'system of qualities in themselves', and even 'property' make no sense except as abbreviations or abstractions. In fact, a mathematical and semantical analysis of the concept of a property will show us shortly that every property – except of course the null property, which is a fiction – is possessed by some individual or other: there are no properties that fail to be paired to any individuals. In particular, substantial properties are properties of substantial individuals.

In other words, an attribute can only be attributed to, or predicated of, some subject or other. A snowball is white: whiteness can be predicated of snowballs and other things but it does not exist separately any more than snowballs failing to "participate" in whiteness – to lapse into Platonic jargon. In other words there are no universals in themselves but only properties that are universal in a given set of individuals, i.e. that are shared by all the members of the set. (More on universals in Sec. 5.3.) All this is rather obvious from the way properties are handled in science and predicates are analyzed in semantics. In both cases every property is represented by a function mapping individuals (or n-tuples of individuals) into statements of the form ⌜Individual (or n-tuple) x is assigned attribute A⌝. Such propositional functions, or proposition-valued functions, will be called *attributes* or *predicates*.

Example 1 The property of being able to read is representable as a function R_1 from the set H of humans into the set P_1 of propositions containing the predicate R_1:

$R_1: H \rightarrow P_1$, where $R_1(b)$, for $b \in H$, abbreviates "b can read", a member of the set P_1.

Example 2 The property of being able to read books can be represented by a function R_2 from the set of couples $\langle x, y \rangle$, with $x \in H$ and $y \in B$, where $H =$ Humans and $B =$ Books:

$R_2: H \times B \to P_2$, where $R_2(b, c)$, for $b \in H$ and $c \in B$, abbreviates "b can read c", a member of P_2.

Example 3 An arbitrary mathematical function f with domain D and range R can be associated the propositional function

$F: D \times R \to P_3$ such that $F(x, y) = (f(x) = y) \in P_3$ for $x \in D$ and $y \in R$.

In ontology we are concerned with properties of entities, i.e. members of the set S of substantial individuals characterized in Ch. 1. A substantial property must then be representable as a propositional function, or predicate, on a domain that somehow includes S. The function will represent the property in general, e.g. age; and its value for a particular individual, the given property of the individual concerned, e.g. its age. But sets other than S are likely to occur as well in the domain of the predicate. For example, mass is representable by a certain real valued function M on the set of quadruples ⟨body, reference frame, time, mass unit⟩. In symbols,

$$\ulcorner M: B \times F \times T \times U_M \times \mathbb{R} \to \text{Propositions containing } M \urcorner$$

represents the mass of entities $B \subset S$, whereas the value $\ulcorner M(b, f, t, u, r) \urcorner$ represents the mass $r \in \mathbb{R}$ of a particular body $b \in B$ relative to a given frame $f \in F$ at time $t \in T$, reckoned or measured in mass units $u \in U_M$ (e.g. grams or pounds).

We generalize the preceding considerations by introducing the following axiom of a methodological nature:

POSTULATE 2.1 Let S be the set of substantial individuals or some subset thereof, and let T to Z be arbitrary nonempty sets, equal to or different from S. Then

(i) any *substantial property in general* is representable as a predicate (or propositional function) of the form

$$A: S \times T \times \ldots \times Z \to \text{Propositions including } A;$$

(ii) any *individual substantial property*, or property of a particular substantial individual $s \in S$, is representable as the value of an attribute at s, i.e. as $A(s, t, \ldots, z)$, where $t \in T, \ldots, z \in Z$.

Every individual property, or property of a particular, is *dichotomic* in the sense that the individual either has it or fails to possess it. This holds not only for qualities and for quantitative properties that take on definite numerical values but also for any quantity the range of which is a family of intervals, as is the case with quantum theoretic dynamical variables. Therefore in an expression of the form $\ulcorner A(s, t, \ldots y) = z \urcorner$ it is irrelevant, for our present purposes, whether z is a single number (or an n-tuple of numbers) or a whole set of numbers (or an n-tuple of numerical intervals). In either case, whether sharp or extended, an individual or a set, z will represent a property of individual s.

Properties are then attributable to, or predicable of, individuals of some kind or other. There is no substantial property apart from entities, let alone prior to them and dwelling in a separate Realm of Forms: a form is a form (property) of a chunk of substance, a universal is a property possessed by all the substantial individuals in a given subset of S. Substrate free forms, or Platonic universals, are just as imaginary as formless substances. But at least the fiction of the bare individual makes mathematical sense while that of a pure form does not. (The barest of individuals is characterizable as the member of some set, but the simplest of forms requires a support or domain, because it is a function.)

An entity possesses certain properties but *is* not a bundle of properties. That entity b possesses properties $P_1, P_2, \ldots, P_n, \ldots$ does not entail that $b = \{P_1, P_2, \ldots, P_n, \ldots\}$. If it did we would have to define properties independently of any individuals, and condone absurd expressions such as '$\{P_1, P_2, \ldots, P_n, \ldots\}$ possesses P_1'.

Example 1 A corpuscle is not a mass-position-velocity compound but an entity possessing definite mass values, position values, and velocity values. (Such a triple is only a schema of such an entity. Recall Vol. 1, Ch. 3, Sec. 3.1.) *Example 2* Having a certain thought is a property of some animals at certain times, and as such is representable by a function on the cartesian product of the set of animals of a certain kind by the set of instants. Put another way: that thinking is a function of certain brains entails that neither thoughts nor brains of certain kinds exist separately. What do exist are brains capable of thinking. (More on the mind-body problem in Vol. 4, Ch. 10.)

The functional construal of properties solves a number of problems, among them the ancient biological conundrum: Which is prior, the organ or the function – e.g. the brain or ideation, feeling, etc.? Answer: neither, for the function is just what the organ does. Take for instance the proposition ⌜The function of organelles of kind A is to synthesize proteins of kind B⌝. This statement can be compressed into ⌜All organelles A synthesize some protein B⌝. The latter does not contain the suspect notion of biological function, which smacks of purposiveness, nor does it state that the synthesis of B is the sole function of an A. The second statement exhibits clearly the idea that the function concerned is a relation between organelles and proteins, namely the relation of synthesizing, which can be construed as a function from the set $A \times B$ of organelle-protein pairs into certain propositions. In this way the quaint idea that there could be a biological function independent of either an organ or its output vanishes.

2.2. *Intrinsic and Mutual, Primary and Secondary*

Some properties, such as radioactivity and intelligence, are inherent properties of individuals. Therefore we can sometimes represent them by unary attributes:

$$Rad: \text{Atoms} \rightarrow \text{Propositions involving } Rad$$

$$Int: \text{People} \rightarrow \text{Propositions involving } Int.$$

We call such properties *intrinsic*. Other properties, such as solubility and performance, are properties of pairs or, in general, n-tuples of substantial individuals. Accordingly we represent them by predicates of rank 2 or higher, such as

$$Sol: \text{Solutes} \times \text{Solvents} \rightarrow \text{Propositions involving } Sol$$

$$Per: \text{People} \times \text{Circumstances} \rightarrow \text{Propositions involving } Per.$$

Such properties will be called *mutual* or *relational*.

All mutual properties must be represented by predicates of rank higher than one but the converse is false. That is, some n-ary predicates, with $n > 1$, represent intrinsic properties. For example, the gross national product, which is an intrinsic property, is representable by a predicate of the form

$$GNP: N \times T \times U \times Q^{+} \rightarrow \text{Propositions involving } GNP,$$

where N is the family of nations, T the set of years, U the set of units of production (e.g. dollars), and Q^{+} the set of positive fractions. In the determination of the n-arity of a property,what counts is the number of sets of substantial individuals occurring in the domain of the corresponding predicate.

A mutual or relational property of an entity may or may not depend causally upon some other individual. For example velocity is a mutual property since it depends on both the moving entity and the reference frame, but it is not caused by the latter. The same holds for distance, duration, frequency, mass, temperature, electric field strength and many other properties in relativistic physics: they are frame-dependent but not causally dependent upon the reference frame, which is assumed to be passive. On the other hand certain mutual properties are environment-dependent. This is the case with the force on a body, the position and momentum distribution of a microsystem, solubility, the frequency of the cricket chirping, the performance of a student, and the rate of production of manufactured goods: in all these cases the environment exerts an influence on the entity of interest. All the phenomenal properties, such as color and perceived loudness, are mutual properties of this kind, i.e. they depend not only upon the object-in-the-environment but also upon the subject or perceiver. More on this presently.

(The way to find out whether a given property is absolute (or invariant or frame-independent) is to investigate its behavior under changes in the reference frame – e.g. under displacements or under rotations. If the corresponding attribute does not change under the given transformation then it is declared to be *invariant* under the latter (or an invariant of the entire group of transformations). This procedure establishes or refutes the hypotheses of relative invariance, i.e. absoluteness with respect to a given group of transformations. Since there is no way of finding out whether a given property is invariant with respect to an arbitrary group of transformations, there is no such thing as absolute invariance (or absoluteness). That is, *all invariance is relative.*)

Since phenomenal properties are subject-dependent they are, a fortiori, frame-dependent. (Indeed, a sentient being may be regarded as a reference frame.) The converse is not true, i.e. not every frame-dependent property is subject-dependent or subjective. For example, the frequency of an oscillation is frame-dependent but not subject-dependent. In case of doubt concerning the objective character of a property, swap subjects (observers) either experimentally or in the theoretical formulas, and see whether the values of the property change. If they do not then the property is objective on top of being relational. Outside psychology all properties, whether intrinsic or mutual, are supposed to be objective, i.e. observer-independent. This alone, and in particular the existence of invariant properties (such as electric charge, entropy, and number of entities), disqualifies any subjectivistic ontology in the eye of science. In other words, a scientific metaphysics must be just as objectivistic as science itself, i.e. thoroughly.

A scientific ontology will not discard phenomenal or secondary properties, such as color and loudness, but will attempt to account for them in terms of nonphenomenal or primary qualities. Whatever the precise account may turn out to be, it must rest on the construal of a secondary property as a mutual property at least one of whose legs is a set of sentient organisms. For example, color is wavelength as perceived by some subject, loudness felt acoustic intensity, and warmth felt heat. No sentient organism, no secondary properties. Hence color predicates are representable as certain functions on the cartesian product of the set of light signals by the set of organisms endowed with sight. The other phenomenal predicates (i.e. predicates representing secondary qualities) are similar.

The physicist assumes that there are sound waves that go unheard and light waves that are unseen. And the psychologist knows that there can be acoustic and visual sensations in the absence of concurrent physical inputs. Therefore it would seem that neither primary qualities are necessary for the corresponding secondary qualities nor conversely. However, there would be no perceptual hallucinations if there were no normal perceptions: psychologists assume tacitly that an organism deprived of every trace of a hearing system experiences no sound hallucinations, and similarly for the other kinds of sensation. Hence sound waves are necessary for hearing – normally as immediate causes, abnormally, as stimuli for past experiences (ontogenetically, phylogenetically, or both). We shall accordingly say that acoustic volume *supports* loudness, luminosity *supports* brightness, and so on, even though not every sensation is caused in an immediate fashion by a physical stimulus endowed with the corresponding supporting ability. This view goes back to Locke (1689 Bk. II, Ch. VIII).

The preceding considerations make no sense to phenomenalism, operationism, and subjectivism: according to these philosophies all properties are secondary. But of course natural science since Galilei recognizes only primary properties, and its progress has partly consisted in shoving secondary qualities to the field of psychology. Note finally that in our view secondary qualities are neither purely objective nor purely subjective, for they are possessed by the subject/object interface rather than by either component separately. This is why a secondary quality is construed as a mutual property of entities and sentient organisms.

We may summarize this subsection in the following methodological axiom, which spells out Postulate 2.1:

POSTULATE 2.2 Let $\{S_i \subseteq S | 1 \leqslant i \leqslant n\}$ be a family of nonempty subsets of substantial individuals not including the null individual. Further, let T to Z be arbitrary nonempty sets [of e.g. units], the same as or different from any of the S_i. Finally, let $\mathbb{R}$ be a set of numbers or of sets of numbers [such as the power set of the real line] and p a natural number. Then

(i) every *qualitative intrinsic property* (quality) of the S_i is representable by attributes of the form

$$A: S_i \times T \times \ldots \times Z \to \text{Propositions containing } A;$$

(ii) every *qualitative mutual property* of the S_i, for $1 \leqslant i \leqslant m \leqslant n$, is representable by attributes of the form

$$A: S_1 \times S_2 \times \ldots \times S_m \times T \times \ldots \times Z \to \text{Propositions containing } A;$$

(iii) every *qualitative phenomenal* (secondary) *property* is a qualitative mutual property representable as a predicate of type (ii) where $T =$ Set of sentient organisms;

(iv) every *quantitative intrinsic property* of the S_i is representable by attributes (magnitudes, quantities, variables) of type (i) where $Z = \mathbb{R}^p$;

(v) every *quantitative mutual property* of the S_i for $1 \leqslant i \leqslant m \leqslant n$ is representable by attributes (magnitudes, quantities, variables) of type (ii) where $Z = \mathbb{R}^p$;

(vi) every *quantitative phenomenal* (secondary) property is a quantitative mutual property representable by a predicate of type (v) with $T =$ Set of sentient organisms;

(vii) every *property of an individual* (or individual property) ${}^i b$ of kind S_i is representable by the value of the corresponding attribute at ${}^i b$, and every *property of an aggregation of m individuals* ${}^1 b \dotplus {}^2 b \dotplus \ldots \dotplus {}^m b$, where ${}^i b \in S_i$, is representable by the value of the corresponding n-ary attribute at $\langle {}^1 b, {}^2 b, \ldots, {}^m b \rangle$.

It might be thought that some if not all of the above distinctions are shallow or arbitrary. In particular, it might be claimed that all individual properties, whether qualitative or quantitative, can be represented by dichotomic (or presence-absence) attributes. Thus both "is married" and "is 25 years old" are dichotomic attributes in the sense that they are either true or false of a given individual. But the point is that, whereas being married is a qualitative property that admits of no degrees, age does come in degrees. What is true is that the progress of science carries with it the quantitation of a number of attributes that had previously been regarded as being inherently dichotomous.

Likewise it might be thought that, since all relations can be reduced to binary relations (Quine, 1954), the n-arity of a predicate is unimportant. This may well be so, but such a logical reduction need not be construed as mirroring anything in reality. Our planet did not cease to revolve around the Sun between Mercury and Saturn the moment ternary relations were shown to be reducible. Besides, the said reduction does not hold for the most useful of all relations, namely functions. Thus a function of two variables is not generally reducible to two functions of one variable each (i.e. two dyadic relations).

Nor does the opposite move work, viz., that of augmenting the degree of predicates on philosophical grounds. This gambit was tried by Helmholtz (1873, p. 260), who held that "each quality or property of a thing is, in reality, nothing else but its capability of exercising certain effects upon other things". Likewise Ducasse (1968) maintained that every property is a causal power to produce some effects. This view rests on the operationist confusion between a property and the way of testing for it. Surely a property that does not affect anything cannot be observed even indirectly and therefore can hardly be said to be possessed by the entity of interest. But if scientific theory construes a given property as an n-ary predicate, our metaphysics has no right to "interpret" it as an $(n+1)$-ary property.

3. THEORY

3.1. *Unarization and Dichotomization*

Thus far we have made only a few rather informal remarks on properties and have analyzed the way they are representable by predicates. In the present section we shall make definite ontological assumptions embodied in a general theory of properties and their scopes, i.e. kinds. Given the variety of properties (intrinsic and mutual, qualitative and quantitative, primary and secondary) it may well be wondered whether the task is feasible. Fortunately we can introduce a remarkable uniformity, and thus pave the way for the search for structure, by using two devices. One is to replace every mutual property by a bunch of intrinsic properties, the other is to concentrate on individual intrinsic properties. The end result is in each case a set of unary dichotomic predicates every one of which is either true or false of a given substantial individual. A couple of examples will show how to proceed in general.

Consider the property of falling. In ordinary knowledge this is represented by a unary predicate. In science, falling is analyzed as a binary predicate F such that 'Fxy' is interpreted as "x falls on y". We can now freeze the second argument, i.e. take it for granted that whatever falls does so on a fixed body b such as our planet. That is, we can form the *pseudo-unary* predicate F_b such that

$$F_b x = x \text{ falls on } b.$$

If now we change the value of the parameter, say to $c \neq b$, we obtain another predicate, namely F_c, such that $F_c y = y$ falls on c. In this way *the single binary predicate F is replaced by an infinite set of unary predicates* F_z. (No return then to the original innocence of ordinary knowledge.) There is of course no economy at all in the unarization procedure: this is just a trick allowing one to speak, though at greater length, of any property of an individual *as if* it were intrinsic. Therefore it has nothing to do with either Bradley's attempt to eliminate relations in favor of monadic predicates or "internal relations" (Bradley, 1893). Nor is it related to the even more quaint claim that a binary relation can be construed as composed of three separate parts: an outgoing arrow, an incoming arrow, and the glue between them (Harary, 1971).

The unarization procedure can be generalized to cover not only mutual properties of n-tuples of individuals but all kinds of complex properties. Consider the statement

> ⌜Body a pulls body b, at time t, relative to reference frame f, with force p in force units u⌝.,

or ⌜$P(a, b, t, f, u, p)$⌝ for short. The 6th rank quantitative predicate P represents the property of pulling, and any of its values represents the individual property of a given thing pulling another with a fixed force relative to some frame at some instant. By focusing on the agent a we manufacture one predicate P_{btfup} for each circumstance $\langle b, t, f, u, p\rangle$, such that

$$P_{btfup}(a) = a \text{ pulls } [b \text{ at time } t \text{ relative to frame } f \text{ with force } p \text{ in units } u].$$

Every one of these predicates may be taken to represent an intrinsic property and moreover an individual one, for it is either possessed or not possessed by the corresponding substantial individual, in this case a. Since the sets of instants, frames, force units, and force values are infinite, the number of such artificial unary predicates is infinite as well. This is a high price to pay and for this reason the transaction is not performed in science. But it is the price to pay if we want a general theory of properties and kinds. Let us then proceed to make

DEFINITION 2.1 Let $S_1, S_2, \ldots, S_n \subseteq S$, with $n > 1$, be n nonempty sets of substantial individuals and let $T_1, T_2, \ldots, T_m$, with $m \geqslant 0$, be m arbitrary nonempty sets. Furthermore let

$$A : S_1 \times S_2 \times \ldots \times S_n \times T_1 \times T_2 \times \ldots \times T_m \to \text{Propositions involving } A$$

be an attribute representing an n-ary substantial mutual property of such individuals. Then for each $1 \leqslant i \leqslant n$ and for each $(n+m-1)$-tuple

$$\langle x_1, x_2, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, t_1, t_2, \ldots, t_m \rangle$$
$$\in S_1 \times S_2 \times \ldots S_{i-1} \times S_{i+1} \times \ldots \times S_n \times T_1 \times T_2 \times \ldots \times T_m,$$

the *i-th unary attributes associated to A* are the functions

$$A x_1 x_2 \ldots x_{i-1} x_{i+1} \ldots x_n t_1 t_2 \ldots t_m(x) =$$
$$=_{df} A(x_1, x_2, \ldots, x_{i-1}, x, x_{i+1}, t_1, t_2, \ldots, t_m).$$

We shall henceforth take this unarization for granted throughout the remainder of this chapter. And we shall also take it for granted that, when dealing with a property, we actually handle an individual property, i.e. a property of some substantial individual. These two conventions allow us to introduce

DEFINITION 2.2 P is a *substantial property* (or a member of the set $\mathbb{P}$ of substantial properties) iff some substantial individuals possess P:

$$P \in \mathbb{P} =_{df} (\exists x)(x \in S \ \& \ x \text{ possesses } P).$$

In what follows we shall investigate the structure of the set $\mathbb{P}$ of substantial properties. That such an investigation is called for is obvious from the result of Sec. 1, namely that although all properties may be representable by attributes, not all attributes represent substantial properties.

3.2. *Basic Assumptions and Conventions*

We proceed to formulate a handful of assumptions and conventions that will allow us to lay bare the structure of the set of all substantial (intrinsic and dichotomic) properties. (The formal treatment from Definition 2.5 on follows closely Bunge and Sangalli (1977).)

The collection of properties of an individual, and that of a set of individuals, deserve special names that will be used frequently in the sequel:

DEFINITION 2.3 Let $T \subseteq S$ be a nonempty set of substantial individuals and $\mathbb{P}$ the set of (unarized dichotomic) substantial properties. Then

(i) the set of (unarized dichotomic) *properties of individual* $x \in T$ is called

$$p(x) = \{P \in \mathbb{P} \mid x \text{ possesses } P\};$$

(ii) the set of (unarized dichotomic) substantial *properties of the* T's is called

$$p(T) = \{P \in \mathbb{P} \mid \text{For all } x \in T, x \text{ possesses } P\}.$$

Consider for a moment the set $p(x)$ of (unarized dichotomic) properties of an entity x. Since every one of these properties is taken at a given time and relative to a given reference frame, when reckoning with all the possible instants and frames we end up with a nondenumerable infinity of (potential) properties of x, no matter how modest this individual may be. (Recall Sec. 3.1.) On the other hand it seems reasonable (and comforting) to assume that the set of substantial properties in general, though extremely numerous, is finite. Thus one thinks of an animal as having different weights, metabolic rates, and ages at different times; yet all these are just different values of only three general properties – weight, metabolic rate, and age at different times and relative to different frames. In any case we stick our neck out and assume

POSTULATE 2.3 The set of general substantial properties is finite, and that of substantial individual (unarized and dichotomized) properties, i.e. $\mathbb{P}$, is nondenumerable.

Next we accept Aristotle's ontological axiom that every individual either possesses a given property (in a given respect at some time) or not:

POSTULATE 2.4 For every $x \in S$ and every $P \in \mathbb{P}$,

$$x \text{ possesses } P \vee \neg x \text{ possesses } P.$$

This axiom must not be mistaken for the excluded middle principle, which concerns predicates. Hence a mathematical intuitionist can accept the former. By logic, whether ordinary or intuitionistic, it follows immediately that no individual has a given property and fails to possess it. That is, there are no inherently contradictory things: contradiction proper is always *de dicto* never *de re*. What there can be is of course mutually opposing properties, such as excitation and inhibition, but these do not illustrate contradiction. Also, by adopting Postulate 2.4 or its corollary, one does not rule out change, in particular the acquisition or loss of a property, because every property occurring in the postulate is taken at a given time under given circumstances. Which leads us to

DEFINITION 2.4 Two properties are *incompatible* over a set $T \subseteq S$ of substantial individuals iff possessing one of them precludes having the other: If $P_1, P_2 \in \mathbb{P}$, then

(i) P_1 is *incompatible* with P_2 over $T =_{df} (x)(x \in T \Rightarrow (x \text{ possesses } P_1 \Rightarrow \neg x \text{ possesses } P_2))$;

(ii) P_1 and P_2 are mutually *compatible* over T iff they are not incompatible over T.

Note that two properties may be compatible yet not both actually possessed by a given entity. For instance honesty and wealth are compatible even though rarely concomitant. (Compatibilities are dispositional properties.)

An interesting case of property incompatibility is that of microsystems. According to the quantum theory a microsystem does not have both a sharp spatial localization and a precise velocity value. (What are compatible, moreover concomitant, are the probability distributions of position and of momentum.) In this theory several properties (the dynamical variables) are represented by operators, some of which fail to commute pairwise. If A_1 and A_2 are operators representing mutually incompatible properties, then $A_1A_2 \neq A_2A_1$ and conversely. When this is the case then, whenever A_1 takes on a precise (eigen)value, its partner A_2 exhibits a whole range of values and conversely. That is, noncommuting variables do not have joint sharp values: they have only distributions of values, each with a given probability. Operationism notwithstanding this – insofar as quantum mechanics is true – is an objective trait of nature not a feature of measurement. In other words the incompatibility between properties represented by noncommuting operators is verified in experiment not caused by the latter (Bunge, 1967b).

We adopt next a *principium individuationis*:

POSTULATE 2.5 No two substantial individuals have exactly the same properties:

$$\text{For all} \quad x, y \in S, \quad \text{if} \quad x \neq y \quad \text{then} \quad p(x) \neq p(y).$$

By contraposition it follows that, if two entities have exactly the same properties, then they are one:

COROLLARY 2.1 For all $x, y \in S$,

$$\text{if} \quad p(x) = p(y), \quad \text{then} \quad x = y.$$

Either of these hypotheses may be called *Leibniz' law* (Leibniz, third and fourth letter to Clarke, in Alexander, Ed., 1956). They take identity seriously without mistaking it for mere similarity: the slightest difference between two entities – such as a difference in relative position with respect to a third entity – results in difference. It is only in the realm of constructs that there can be several identical copies of one and the same object, as with the cases of $a + a$ and $A \times A$. But even here one may argue that there is a single object taken repeatedly.

It is usually maintained that the concept of identity occurring in Leibniz' law is that of *contingent identity* as opposed to formal identity. Thus a living brain is said to be contingently identical with the interior of the skull of a living animal, and gene duplication contingently identical with the uncoiling of the double helix. I do not understand the alleged difference between contingent and formal identity: since both concepts have exactly the same properties they must be identical and must fall under the theory of identity.

Because the relation of identity is reflexive, we retrieve a famous statement sometimes regarded as a basic principle of both logic and ontology, at other times decried as absurd or, even worse, as trivial, namely

COROLLARY 2.2 Every entity is identical to itself (i.e. self identical).

Some philosophers have held that this statement is incompatible with change. Nothing of the sort. If an entity changes then it becomes a different entity – or an entity in a different state – even though we may go on calling it by the same name. All that Corollary 2.2 asserts is that every entity preserves whatever "identity" it may have – until it loses it and acquires another. (Caution: in the last sentence 'identity' signifies "collection of peculiarities".)

Note that identity, therefore also difference, are relations, hence concepts not facts. Even though it is true that a dog is not the same as a star, there is no connection or coupling between the two – i.e. "$\neq$" mirrors no such relationship or link. In other words both identity and difference are *de dicto*. This is one of the crucial places at which naive realism (i.e. the fact-thought isomorphism thesis) fails.

One last caution concerning identity: this concept must not – *pace* Strawson (1959) – be mistaken for the pragmatic concept of *identification*. One may by mistake identify two things that are actually different; or one may "identify" an object *qua* member of a certain class, thereby failing to use the concept of identity. For this reason we cannot adopt the operationist claim that "the relation of identity or distinctness has no other meaning than that two objects have been identified or distinguished" (Yessenin-Volpin, 1970, p. 7). Whether or not an identity statement has been put to the test, it should be meaningful to begin with. (For the relation between meaning and testability see Vol. 2, Ch. 7, Sec. 5.1.)

Finally we reach a central convention in our theory:

DEFINITION 2.5 The *scope* of a substantial property is the collection of entities possessing it. In other words, the scope $\mathcal{S}$ is the function $\mathcal{S}\colon \mathbb{P} \to 2^S$ from the set of all substantial properties to the set of all the subsets of substantial individuals, such that "$x \in \mathcal{S}(P)$", for $x \in S$, is interpreted as "*Individual x possesses property P*".

For example, the scope of the mass property is the set of all bodies, and that of the social mobility property is the set of all human societies.

A first, surely modest use of $\mathcal{S}$ is in defining the fiction called *null property*: N is a null property iff $\mathcal{S}(N) = \varnothing$.

Note the difference between the scope of a property and the extension of a predicate or attribute (cf. Vol. 2, Ch. 9, Sec. 1). For one thing $\mathcal{S}$ is defined on properties not on predicates; for another $\mathcal{S}$ is not defined for negation and disjunction.

Note also that some of the ideas handled in the preceding can be reformulated in terms of scopes. For example the Definition 2.4(i) of property incompatibility becomes

$$\begin{aligned}&P_1 \text{ and } P_2 \text{ are } \textit{mutually incompatible} \text{ over}\\ &T =_{df} \mathcal{S}(P_1) \cap \mathcal{S}(P_2) \cap T = \varnothing.\end{aligned}$$

Given a property other than a null property, there is always a nonempty set of individuals that possess it since, by Definition 2.2, a substantial property is one possessed by at least one substantial individual. See Figure 2.1. But the converse is false: an arbitrary collection of objects need not share a given property. When they do, and no object outside the collection has the property of interest, the set is called a class:

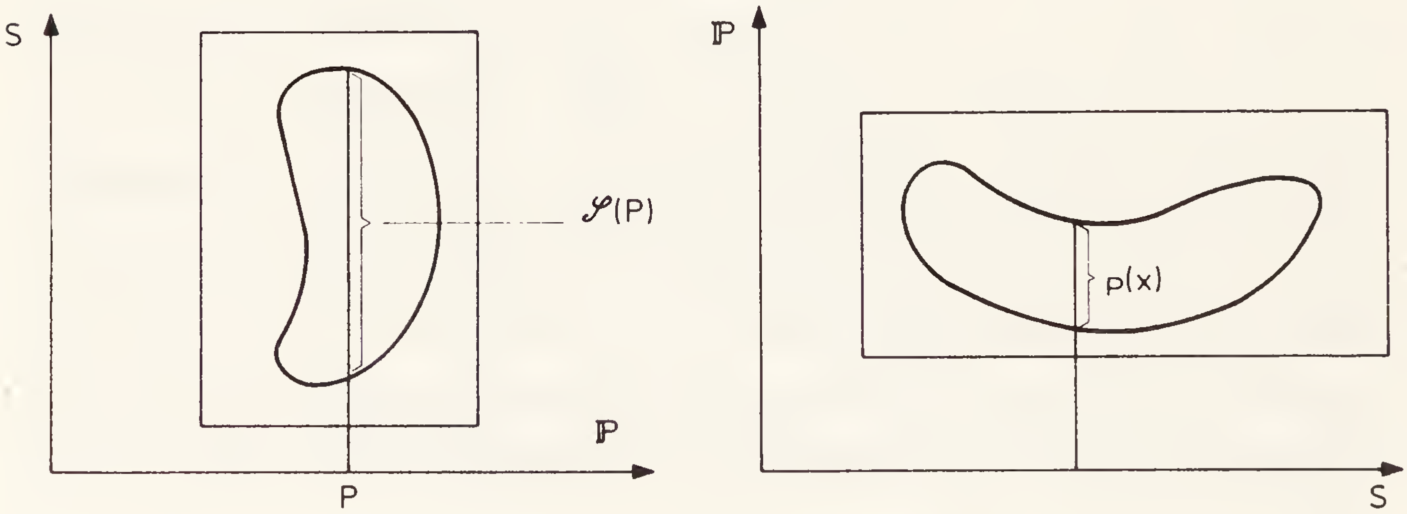


Fig. 2.1. Two complementary concepts: the scope of a property and the set of properties of an individual.

DEFINITION 2.6 A subset X of S is called a *class* (of substantial individuals) iff there exists a substantial property $P \in \mathbb{P}$ such that $X = \mathcal{S}(P) \in 2^S$.

Since all substantial individuals can associate (Ch. 1), S forms a class – in fact the largest of all classes studied by ontology, i.e. the universe of discourse of the latter.

Our next assumption is

POSTULATE 2.6 The intersection of any two classes is a class: For any two compatible substantial properties $P, Q \in \mathbb{P}$, there is at least one other property $R \in \mathbb{P}$ such that $\mathscr{S}(R) = \mathscr{S}(P) \cap \mathscr{S}(Q)$.

Note that the third property determined by any two given properties need not be unique. In fact there might be a fourth property with the same scope as R. Secondly, if P and Q are mutually incompatible over some domain of individuals, then they determine an empty class. Thirdly, the above assumption is not necessary for attributes or predicates because, if A and B are attributes with overlapping domains, then their conjunction $A \& B$ exists on their overlap and it determines a set, namely the extension of the composite predicate $A \& B$. But we still do not know what the conjunction of properties might be, so the above postulate is nontrivial. Fourthly, we are not asserting that the union of two classes is a class, for this would be definitely false in the particular case when the classes are natural kinds: think of the set (not the class) equal to the union of the class of clouds and the class of flies.

So much for properties in general. Let us now study a particular class of properties, namely laws.

3.3. *Laws as Properties*

So far as Postulate 2.6 is concerned, all the classes could be pairwise disjoint. Our final assumption will deny this possibility, i.e. it will assert that some pairs of classes overlap and, moreover, that some are contained in others. In other words, we shall assume that all entities satisfy some laws – or, what amounts to the same, that every substantial property is lawfully related to some other substantial property. To formulate this idea with precision we need

DEFINITION 2.7 If $P, Q \in \mathbb{P}$ are substantial properties, then the statement $\ulcorner \mathscr{S}(P) \subseteq \mathscr{S}(Q) \urcorner$ or its converse or, equivalently,

> $(x)Lx$ with $Lx = \ulcorner x$ possesses $P \Rightarrow x$ possesses $Q \urcorner$ or its converse

is called a *law statement* relating P to Q.

If the scope of a property is included in that of another, the two properties are said to be *lawfully related*. A property not so related to any other property is said to be *stray* or *lawless*.

This notion of a law will be considerably refined in Ch. 3, Sec. 2.3. Note that we have not defined laws, or objective patterns, but law statements, or conceptual reconstructions of objective patterns. The relation between the two categories is an instance of the property-attribute relation discussed in Sec. 1.2.

It might be objected that there are single property laws, such as the law of inertia, which can be abbreviated to "$V = \text{const.}$" This would be mistaken, for the preceding formula is a poor rendering of the correct statement, which involves further variables. Firstly there is the property of being a free body (predicate B). Secondly, there is a tacit reference to a frame (predicate F) as well as to time (T). Indeed, the complete formulation is: "If x is a free body (B) and y a reference frame (F) and t an instant of time (T), then the velocity of x relative to y is constant for all t in T". The function V represents then a mutual property (of body and frame) and the last statement makes no sense without the antecedent, which contains the predicates B and F. True, time (T) looks external to things – common to all hence the property of none. But it actually summarizes the state of the reference frame. (See Ch. 5, Sec. 3.2.) In fact t can be replaced by the phase of some process occurring in a frame $y \in F$. (For example t could be the angular position of the pointer of a chronometer attached to y, or the mass of water flowing in a clepsydra.) This renders it a property of entities. In any event the initial simplicity of the law of inertia was deceptive.

We can now formulate the hypothesis that there are no stray or lawless properties:

POSTULATE 2.7 Every substantial property is lawfully related to some other substantial property. I.e., if $P \in \mathbb{P}$, then there exists a $Q \in \mathbb{P}$ such that either $\mathscr{S}(P) \subseteq \mathscr{S}(Q)$ or $\mathscr{S}(Q) \subseteq \mathscr{S}(P)$.

This ontological principle underlies all science and all technology. (Recall Introduction, Sec. 7, principle *M8*.)

Note that, since laws interrelate substantial properties and the latter are properties, *laws themselves are properties of entities*. So much so that they can be expressed by formulas of the basic form $\ulcorner (x)Lx \urcorner$. Thus it is a property of a free body that it satisfies the law of inertia. Like every other property, it is represented by a predicate analyzable in the manner prescribed by Postulate 2.2. However, laws are properties *lato sensu*, not in the strict sense we have been considering heretofore. For, if every law were to belong to $\mathbb{P}$, then by Postulate 2.7 it would occur in some other law – and so on *ad infinitum*. But this would contradict the Postulate 2.3 that $\mathbb{P}$ is finite. Hence when claiming that laws are properties we use the term 'property' in a broader sense than when we state that laws relate properties. More in a while.

Postulate 2.7 is our present version of the *principle of lawfulness*. (For a discussion of the latter see Bunge (1959).) It will play a decisive role in Ch. 3, Sec. 3.3, in helping us define the notion of a natural kind as distinct from a mere class. And it provides the following propertyhood

CRITERION 2.1 P is a property iff P occurs in at least one law.

By this criterion predicates such as "non-smoker", "fast or thinking", and "green before AD 2000 and blue thereafter" (the infamous "grue") do not represent any substantial properties. This is of course in agreement with Sec. 1. Moreover we are now in a better position to distinguish an arbitrary function from one representing a substantial property:

CRITERION 2.2 A function F represents a property of substantial individuals of class $T \subseteq S$ iff

(i) T^n, with $n \geqslant 1$, occurs in the domain of F;

(ii) F occurs in at least one law statement referring to T.

By this criterion "differentiable" (in the mathematical sense) does not represent a substantial property because its domain is the set of functions. On the other hand "differentiable function" can be made to represent properties that happen to be smooth.

Finally we introduce a useful convention:

DEFINITION 2.8 The totality of laws possessed by substantial individuals is denoted by $\mathbb{L}$.

Remark 1 The ontological concept of a law can be illustrated. On the other hand it is not ostensive: it is impossible to point to a law – as different from a sentence expressing a law statement conceptualizing a law. Furthermore nobody could possibly exhibit or even mention the set of all laws. (At most one could try to list the known law statements in a given field of research – or rather a subset of standard law statements with neglect of their innumerable instances.) Thus $\mathbb{L}$ is a nonconstructive set and is, if anything, even more metaphysical than its members. It is one of those ontological notions used, though not elucidated, in the everyday metascientific discourse of scientists. *Remark 2* Since laws are properties in a broad sense, we can form the totality $\mathbb{P} \cup \mathbb{L}$ of properties *lato sensu*, where $\mathbb{P}$ is the set of properties *sensu stricto* (those occurring in the members of $\mathbb{L}$). *Remark 3* Because laws are properties, they can be assigned scopes – provided we agree to introduce a new scope function defined on $\mathbb{P} \cup \mathbb{L}$, not just on $\mathbb{P}$. The scope of a law would then be the set of entities possessing ("obeying") the law. This set is the same as the reference class of the law statement representing the objective law – not however the same as its extension; which is the set of entities for which the law statement holds exactly. *Remark 4* Another consequence of construing laws as properties (in a broad sense) is that the scope of a law may include or be included in that of some other property, in particular that of another law. Hence the set $\mathbb{L}$ of laws, far from being amorphous, is partially ordered in respect of breadth of scope. However, in the technical developments to follow we shall keep to the narrow construal of 'property'.

So much for our main assumptions and definitions. Let us now proceed to make the most of them.

3.4. *Precedence and Conjunction of Properties*

The concept of a law introduced by Definition 2.7 and occurring in Postulate 2.7 suggests the formation of the following preordering relation:

DEFINITION 2.9 If P and Q are substantial properties (i.e. members of $\mathbb{P}$) then *P precedes Q* iff P is more common than Q, i.e.

$$P \preccurlyeq Q =_{df} \mathscr{S}(Q) \subseteq \mathscr{S}(P).$$

Since in turn inclusion is defined in terms of implication, the previous definition is equivalent to

DEFINITION 2.10 If P and Q are substantial properties, then *P precedes Q* iff P is necessary for Q, i.e.

$$P \preccurlyeq Q =_{df} (x)(x \in S \Rightarrow (x \text{ possesses } Q \Rightarrow x \text{ possesses } P)).$$

In other words, "precedes", "is more common than", and "is necessary for" are coextensive. Equivalently: "follows", "is rarer than", and "is sufficient for" are coextensive. Example: the property of thinking follows that of being alive, which in turn follows that of containing genetic material.

Warning: the preceding definitions should not be read in terms of attributes or predicates, if only for the following reason. Let A and B be two attributes with the same referents and the same rank, and assume that B is necessary for A, i.e. $A \Rightarrow B$. Then by Craig's interpolation theorem there exists at least one other attribute C with the same referents and the same rank that mediates between A and B, i.e. such that $A \Rightarrow C \,\&\, C \Rightarrow B$. This, which is a theorem in logic, does not guarantee the existence of a substantial property represented by the interpolated predicate. Such an existence hypothesis, if made, would give us a different theory of properties.

The precedence relation is reflexive and transitive, i.e. it is a preorder relation. In fact it is not antisymmetric, i.e. $P \preccurlyeq Q$ and $Q \preccurlyeq P$ does not imply $P = Q$. Indeed different properties may have the same scope – just as two attributes can have the same extension without being identical. In order to obtain an order and thus uncover richer structures we must shift attention from properties to classes (or scopes of properties). To begin with we make

DEFINITION 2.11 If P and Q are any two substantial properties, then they are *concomitant* iff they have the same scope:

$$P \sim Q =_{df} \mathscr{S}(P) = \mathscr{S}(Q).$$

In other words, two properties are concomitant if the corresponding attributes are equivalent. Or also:

$$P \sim Q =_{df} (x)(x \in S \Rightarrow (x \text{ possesses } P \Leftrightarrow x \text{ possesses } Q)).$$

The concomitance of properties is what Hume called 'constant conjunction of properties' (Hume, 1739–40). In our ontology, by virtue of Definition 2.7, *the concomitance of properties is not coincidental but lawful*; or, if preferred, it is nomically necessary rather than accidental.

The relation $\sim$ of concomitance is important because it is an equivalence relation on $\mathbb{P}$, i.e. it splits $\mathbb{P}$ into disjoint sets (equivalence classes), namely of those properties that happen to come together. To investigate the structure of the resulting sets we need some further notation:

DEFINITION 2.12 Let $\mathbb{P}$ be the collection of substantial properties and $\sim$ the relation of property concomitance defined on $\mathbb{P}$. Then for any $P \in \mathbb{P}$:

(i) the *totality of concomitants* of P: $[P] =_{df} \{Q \in \mathbb{P} \mid Q \sim P\}$;

(ii) the *family of sets of concomitants*: $[\mathbb{P}] =_{df} \{[P] \mid P \in \mathbb{P}\}$;

(iii) the *scope of sets of concomitants*: $[\mathscr{S}]: [\mathbb{P}] \to \mathscr{P}(S)$.

Clearly, for any substantial property P, $[\mathscr{S}]([P]) = \mathscr{S}(P)$ is the class of entities that possess P and all the other properties concomitant with P.

The preorder relation $\leqslant$ on $\mathbb{P}$ induces a partial order relation $[\leqslant]$ on $[\mathbb{P}]$ according to

DEFINITION 2.13 Let $[\mathbb{P}]$ be the collection of all equivalence classes of properties. Then, for any two $[P], [Q] \in [\mathbb{P}]$,

$$[P][\leqslant][Q] =_{df} \mathscr{S}[Q] \subseteq \mathscr{S}[P].$$

This new precedence relation has all the formal properties of the inclusion relation. It enables us to compare not properties but their scopes – thus obliterating any difference among properties with the same scope. Hence $\langle [\mathbb{P}], [\leqslant] \rangle$ is a partially ordered set. Actually $[\mathbb{P}]$ has a richer structure, as shown by

THEOREM 2.1 The family of sets of concomitant properties has the sup-semilattice structure, where the supremum of $[P]$ and $[Q]$ is defined as

$$[P] \sqcup [Q] =_{df} \{R \in \mathbb{P} \mid \mathscr{S}(R) = \mathscr{S}(P) \cap \mathscr{S}(Q)\}.$$

Proof It is easily verified that $[P] \sqcup [Q]$ is a set of concomitants and the least (in the sense of $[\leqslant]$) upper bound of $\{[P], [Q]\}$, i.e. $[P], [Q][\leqslant]([P] \sqcup [Q])$.

COROLLARY 2.1 If a substantial property R belongs to the supremum of two sets $[P]$ and $[Q]$ of concomitant properties, i.e. $R \in [P] \sqcup [Q]$, then an entity possesses R just in case it possesses both P and Q: $(x)(x \in S \Rightarrow (x \text{ possesses } R \Leftrightarrow x \text{ possesses } P \,\&\, x \text{ possesses } Q))$.

We shall shorten "x possesses P & x possesses Q" to "x possesses $P \wedge Q$", where the complex property $P \wedge Q$ is characterized, albeit not defined, by

$$\mathscr{S}(P \wedge Q) = \mathscr{S}(P) \cap \mathscr{S}(Q).$$

$P \wedge Q$ will be called the *conjunction* of P and Q, or the *complex property* composed of P and Q. This is not a definition proper because equality of scope (which is all we are requiring) is necessary but not sufficient for identity of properties. (Caution: property conjunction should not be mistaken for property concomitance or "constant conjunction".)

Properties then may be simple (or basic) or complex (or derivative). For example in contemporary physics mass, charge, spin and strangeness are regarded as basic (simple) properties in the sense that they are not reducible to any other properties. On the other hand weight is a derivative (complex) property, as it consists in an entity's having a mass *and* being placed in a gravitational field. That is, weight is the conjunction of two properties.

Only a scientific theory can say whether a given property is basic or derivative. But such a dichotomy holds only for the proper subset $\mathbb{T}$ of $\mathbb{P}$ represented in the theory, so the most one can venture is to say that a certain property P is simple, or else complex, *in* $\mathbb{T}$. Alternative theories dealing with the same properties may decide differently. For example energy and momentum, which are simple in certain theories, are treated as complex in others. Therefore we cannot pass final judgment on the degree of complexity of a property on the basis of an analysis of the corresponding attributes. Nevertheless such indications, ambiguous and fickle as they are, are the only ones we have.

It seems reasonable to assume that, even though we often find it difficult if not impossible to ascertain whether a given property is basic, reality has no such problem and builds complex properties out of simple ones – or rather complex entities out of simpler ones. It is therefore worth while to state that assumption in exact terms. But to this end we first have to introduce an exact concept of a basic property. And this in turn calls for the explicit introduction of the concept of a universal property, i.e. one possessed by all entities. More precisely, we form the equivalence class of all universal properties and call it

$$[U] =_{df} \{P \in \mathbb{P} \mid \text{For all } x \in S : x \text{ possesses } P\}.$$

Clearly, $[U]$ precedes every other set of concomitant properties in the sup semilattice $\langle [\mathbb{P}], \sqcup \rangle$, hence it is the smallest element of it, i.e. its zero. In other words $[U]$ is the root of the tree of (sets of concomitant) properties: see Figure 2.2. Note that there is no top element formed by the properties that no entity possesses because, by Definition 2.2, a substantial property is possessed by at least one entity.

We are now ready to elucidate the notion of a basic (or generating) property. We stipulate it to be a property occurring as a conjunct in some other (complex) property. More exactly, we adopt

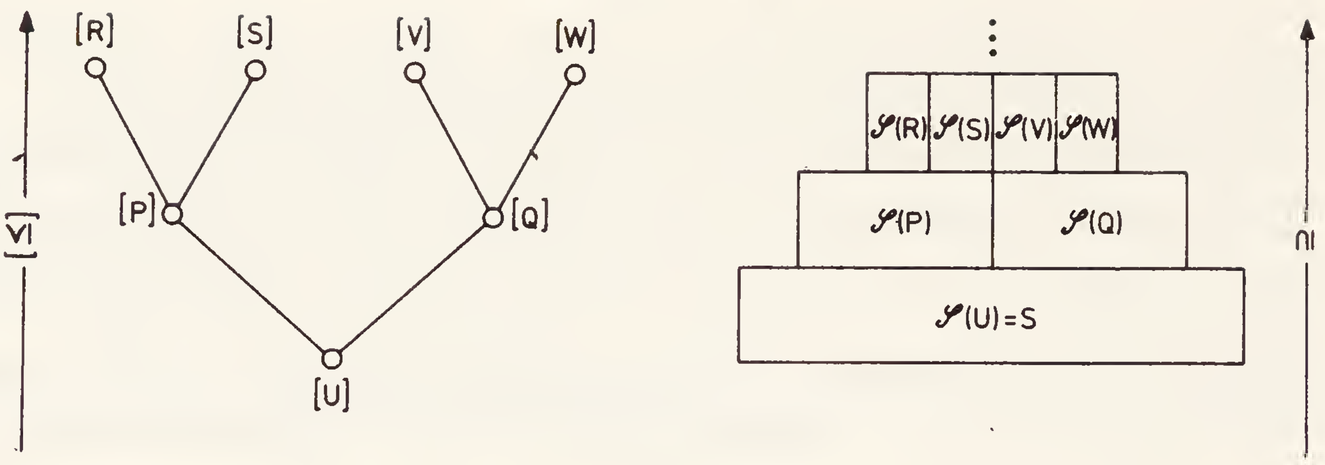


Fig. 2.2. The tree of (sets of concomitant) properties and the pyramid of classes of entities with the same properties. The more basic a property the more common it is.

DEFINITION 2.14 A subset $\mathbb{B}$ of $\mathbb{P}$ is called a set of *independent generators* of $\mathbb{P}$ iff

(i) for every $P \in \mathbb{P}$ there exist $B_1, B_2, \ldots, B_n \in \mathbb{B}$, $n \geqslant 1$, such that $P = B_1 \wedge B_2 \wedge \ldots \wedge B_n$;

(ii) for all $n > 1$ and all $B_1, B_2, \ldots, B_n \in \mathbb{B}$, if $B_1 = B_2 \wedge \ldots \wedge B_n$ then either $B_i = B_1$ or $B_i = U$ (a universal property) for $i = 2, 3, \ldots, n$.

Every set $\mathbb{B}$ of independent generators splits into a set of universal properties and another of *atomic* properties, i.e. properties which are basic (or generating) but not universal:

$$\mathbb{B} \subseteq At(\mathbb{P}) \cup [U].$$

If there are such atomic properties then this is a trait of reality not a mathematical "fact". And if we do assume their existence then we must suppose that they are infinitely many. In fact a finitely generated semi-lattice is finite but, by Postulate 2.3, $\mathbb{P}$ is infinite. Therefore the assumption that there are basic properties that determine all others reads:

POSTULATE 2.8 The set of atoms of $\mathbb{P}$ is infinite and it generates $\mathbb{P}$.

An immediate consequence is that each property is of either of the following types:

(i) *basic* properties (i.e. members of some $\mathbb{B} \subseteq At(\mathbb{P}) \cup [U]$);

(ii) properties that are *finite conjunctions* of two or more basic properties;

(iii) properties that are *infinite conjunctions* of basic properties (in the sense that their scopes are infinite intersections).

Infinite conjunctions of properties are not ghostly or even exceptional. Thus the trajectory of a particle relative to some reference frame may be construed as the conjunction of its (nondenumerably many) successive positions.

The epistemological consequence of Postulate 2.8 is this. For any class $T \subset S$ of entities, in order to know all the properties in $p(T)$ it suffices to know the atomic properties of the T's, i.e. the atoms of $\mathbb{P}$ belonging to $p(T)$.

We close this section with an algebraic note showing that the collection of substantial individuals, far from being an amorphous set, has a definite structure. Consider the family $\mathbb{K}$ of all classes of substantial individuals. Since by Postulate 2.6 the intersection of any two classes is a class, $\mathbb{K}$ has the semilattice structure. In other words, we have proved

THEOREM 2.2 The system $\langle \mathbb{K}, \cap \rangle$, where

$$\mathbb{K} = \{\mathscr{S}(P) \in 2^S \mid P \in \mathbb{P}\}$$

is the family of all classes of substantial individuals, is an inf-semilattice.

We shall have much more to say on the algebra of classes of substantial individuals in Ch. 3, Sec. 3.

3.5. *Similarity*

Knowing that two given concrete objects are different is not enough: we should know what their degree of difference is – or, which amounts to the same, in which respects they are the same. To be the same in some respects is to be similar. Now, similarity can be weak or strong, superficial or deep, according as the overlap in properties is slight or large. Most assertions of identity with reference to concrete objects are actually statements of strong similarity (or equality) not of strict identity. For example the statement that all the atoms of a given kind in a certain state are identical means that they are the same barring certain extrinsic differences – e.g. in position relative to other things.

We must therefore investigate the concept of degree of similarity and its dual, degree of dissimilarity. We start with the convention that the similarity between two things is the collection of their shared properties. More precisely, we make

DEFINITION 2.15 Let $\sigma: S \times S \to 2^{\mathbb{P}}$ be the function such that $\sigma(x, y) = p(x) \cap p(y)$ for any $x, y \in S$. Then

(i) $\sigma(x, y)$ is called the *similarity* between x and y, and

(ii) two entities are said to be *similar* ($\sim$) iff their similarity is not nought:

$$\text{If } x, y \in S \quad \text{then} \quad x \sim y =_{df} \sigma(x, y) \neq \varnothing .$$

Note that $\sim$ is a similarity relation, not an equivalence relation, because it is reflexive and symmetric but not transitive. (One entity may resemble another and the latter a third while the former does not resemble the third. Just think of face similarity.)

By Postulate 1.1 in Ch. 1, Sec. 1.2, all substantial individuals have the property of associating. (Besides, we know from physics that all entities have energy, are capable of moving, etc.) Hence

THEOREM 2.3 All entities are similar [or equal in some respect]: If $x, y \in S$, then $x \sim y$.

In other words, no matter how dissimilar two things may be in most respects, they share some property or other – in fact a number of properties. This is one of the reasons that all things can be studied scientifically (with the help of the scientific method) – the other reason being that they are all studied by beings equipped with similar neurosystems. Caution: this is not to conclude, with mechanism and animism alike, that all things have the same properties, only in varying degrees – as Whitehead (1929) and Teilhard de Chardin (1964) believed.

If, for the sake of expediency rather than truth and depth, we assign all properties the same weight, then we can define a quantitative concept of degree of similarity provided we restrict ourselves to the subset of the known properties of a thing (to avoid division by infinity):

DEFINITION 2.16 The *degree of similarity* between two entities $x, y \in S$ relative to a finite subset $\mathbb{B}$ of $\mathbb{P}$ [such as, e.g., the genetic similarity or the cultural similarity of two human populations] is

$$s(x, y) = \frac{|\sigma(x, y) \cap \mathbb{B}|}{|(p(x) \cup p(y)) \cap \mathbb{B}|}.$$

This is none other than the simple matching coefficient used in numerical taxonomy to estimate the distance between biological species (Sneath and Sokal, 1973). Because $\mathbb{B}$ is a finite set, $s(x, y)$ can be zero or it can be equal to unity. For example, the degree of genetic similarity between any two cells of a single organism equals 1. But by taking $\mathbb{B}$ to be a big enough set of properties we obtain a smaller value of s.

The above concepts of similarity can be generalized to whole sets of entities, namely as follows. Let $x_1, x_2, \ldots, x_n$ be the members of a finite set T of entities. Then the *similarity among the members of* T is defined as

$$\sigma(x_1, x_2, \ldots, x_n) = \bigcap_{x_j \in T} p(x_j)$$

and the corresponding degree of similarity as

$$s(x_1, x_2, \ldots, x_n) = \frac{|\sigma(x_1, x_2, \ldots, x_n)|}{\left|\bigcup_{x_j \in T} p(x_j)\right|}.$$

However, we shall not employ the last two concepts.

The notion of dissimilarity is mathematically more interesting than its dual. We define it thus:

DEFINITION 2.17 Let $\delta: S \times S \to 2^{\mathbb{P}}$ be the function such that

$$\delta(x, y) = p(x) \mathbin{\Delta} p(y) \quad \text{for any} \quad x, y \in S,$$

where 'Δ' designates the symmetric or Boolean difference. $\delta(x, y)$ is called the *dissimilarity* between x and y.

By Definitions 2.15 and 2.17, the relation between δ and σ is

$$\delta(x, y) = p(x) \cup p(y) - \sigma(x, y).$$

The function δ is a set valued distance function, hence $\delta(x, y)$ should be a good measure of the "distance" between x and y in the space of properties. More precisely, we have

THEOREM 2.4 The dissimilarity function δ is a set valued distance function in a one dimensional space: for all $x, y, z \in S$,

(i) $\varnothing \subseteq \delta(x, y) \subset \mathbb{P}$;
(ii) δ is symmetric: $\delta(x, y) = \delta(y, x)$;
(iii) $\delta(x, y) = \varnothing$ iff $x = y$;
(iv) $\delta(x, y) \mathbin{\Delta} \delta(y, z) = \delta(x, z)$.

Proof The first part is obvious from the definition. (ii) is a consequence of the symmetry of Δ. (iii) follows from $p(x) \mathbin{\Delta} p(x) = \varnothing$. Conversely, by

Postulate 2.5 if $p(x)\,\Delta p(y)=\varnothing$ then $x=y$. (iv) follows upon replacing the LHS of (iv) by the definiens:

$$(p(x)\,\Delta\, p(y))\,\Delta\,(p(y)\,\Delta\, p(z))=p(x)\,\Delta\, p(y)\,\Delta\, p(y)\,\Delta\, p(z)=$$
$$=p(x)\,\Delta\, p(z).$$

By taking the cardinality of the dissimilarity and normalizing it to unity we obtain a quantitative concept of dissimilarity:

DEFINITION 2.18 Take a finite subset of the set of all the properties of two entities $x, y \in S$. Then their *degree of dissimilarity* is

$$d(x, y)=\frac{|\delta(x, y)|}{|p(x)\cup p(y)|}.$$

The values of d, just as those of the degree of similarity s, are bounded by 0 and 1. It can be shown that d is a metric, hence a good measure of the distance between entities in respect of their properties. (Needless to say, there is no relation, other than a mathematical similarity, between this distance and a spatial distance.) That is, one can prove

THEOREM 2.5 The structure $\langle S, d\rangle$, where d is the degree of dissimilarity, is a metric space.

Extreme difference is sometimes called *opposition*. This concept is elucidated by

DEFINITION 2.19 Two entities are *opposite* to one another iff their similarity is nil.

But, according to Theorem 2.3, all things share some properties. Hence we have

THEOREM 2.6 There are no opposite entities.

Opposition, when it exists, is relative to a proper subset of properties. Moreover only "positive" properties should count: the mere absence of a property is not to be equated with the opposite of that property. For instance dark is not the opposite of illuminated, as the former is nothing but the absence of illumination. Real opposition, when it exists, is active and consists in having contrary effects. Moreover real opposition does not exist between utterly different entities. For example two animals of the same species may fight over the same pool of water, or the same prey, or the same female, precisely because they have similar drives.

This point deserves a few more examples. In an electric circuit the self-inductance initially opposes the action of the impressed electromotive force, so that the two aspects are mutually opposed in their effects on the overall current. Likewise if a substance A promotes the synthesis of a protein while A' inhibits it, then A and A' may rightfully be regarded as opposites with regard to protein synthesis no matter how much they resemble in other respects – and they do since both can act on the given protein. Likewise in a mammal the sympathetic and the parasympathetic systems oppose one another in the sense that they check each other: when one of them excites a given organ the other inhibits the latter. But even this opposition is not complete, as there is also some cooperation between the two systems. And in any case this interaction ensues in homoeostasis or equilibrium not in qualitative change (such as the breakdown of the main system) – whence it does not illustrate dialectics.

In sum, although *some* pairs of things do have parts and features that are mutually opposed in *limited* respects, no thing is composed of absolute opposites. Consequently the thesis of dialectical ontologies, that *every* thing is a synthesis or unity of opposites (in what respects?) is, if intelligible at all, false. Moreover the assumption that there are absolute opposites contradicts the materialist thesis, incorporated into our system, that all things – no matter how different – share some properties precisely by virtue of being entities. Besides, the thesis that all things are unities of opposites entails the infinite divisibility of every single entity into two mutually opposed entities, since every part should in turn be a synthesis of opposites – and there is no empirical evidence for this consequence. The most we can assert is that there are relative opposites, i.e. things with features that oppose one another in their effects. But this timid thesis can hardly serve as a cornerstone of a philosophical cosmology. (More on dialectics in Bunge, 1975.)

The upshot of this subsection is this. Although by Postulate 2.5 every entity is unique in some respect(s), all entities are similar in some other respect(s). Shall we then say that all entities are alike and aren't? Not at all: 'alike' (or 'similar') is a syncategorematic term or, if preferred, a vague notion that should be qualified by the expression 'in the respect(s) P'. The complete expression 'x and y are alike (similar) in the respect(s) P' is no longer vague, hence it does not lead us to contradiction. It is moreover compatible with 'x and y are dissimilar in the respect(s) Q, where $Q \neq P$'. In other words 'similar' is not an attribute word on the same footing as 'alive': it is an incomplete or partial attribute word – though not more so than 'blue' or 'heavy'. Hence the need for replacing wherever possible the relation $\sim$ of similarity by the function σ or by the function d.

3.6. *Indiscernibility*

We hold that there are no two identical entities – yet it is common experience that some things are indiscernible or indistinguishable. There is no contradiction here, as two different concepts of difference are involved: the *ontological* concept of (objective) difference on the one hand, and the *epistemological* (or pragmatic or psychological) concept of (subjective) differentiability, or discernibility, or ability of somebody to distinguish empirically, e.g. by observation. Two entities may be indistinguishable to a given subject equipped with certain means of observation but may become discernible to the same subject using improved equipment. (The concept of resolving power of a microscope or of a telescope presupposes the difference between objective difference and perceived difference.)

We must therefore distinguish between factual difference and empirical difference or discernibility. This will require roping in a notion that, strictly speaking, we have no right to use systematically since it will be introduced in Volume 4, Ch. 10, namely that of a cognitive subject. Therefore we shall not number the definitions that follow.

DEFINITION. Let $\mathbb{P}_s \subset \mathbb{P}$ be a proper subset of the totality of substantial properties, and s a cognitive subject in a given state and equipped with certain cognitive means (instruments, theories, etc.). Then for $x, y \in S$, x and y are *indiscernible* (indistinguishable) *to* s iff x and y are the same in every respect $P \in \mathbb{P}_s$ known to s. Notation: $x \sim_s y$.

The terms 'indiscernibility' and 'indistinguishability' have been misused at several critical points. *Example 1* Leibniz himself stated his principle (our Postulate 2.5) in an equivocal manner, namely by writing that "il n'y a point dans la nature deux êtres réels absolus indiscernables", and calling this the *principium identitatis indiscernibilius*. As Russell (1900) saw, all Leibniz intended to affirm is "that any two substances [things] differ as to their predicates [properties]". Leibniz cannot have intended to state that if two things are indiscernible then they are identical, for he was well aware of the marvels of microscopy.

Example 2 The logical principle of substitutivity is sometimes called the *principle of the indiscernibility of identicals* – another obvious mistake. Consider: the principle holds that if b has P, and if $b = c$, then P is true of c as well. It has become fashionable to claim that this principle fails in the so-called "intensional contexts" and thus doubts are cast on the truth of Leibniz' ontological principle (our Postulate 2.5). For example, it is clear that "Smith doubts that $x = y$" does not entail "Smith doubts that $x = x$". The rejoinder is that, if Smith doubts that $x = y$, then he will be well advised to abstain from applying it the principle of substitutivity. And, if a principle is not applied, it runs no risks. Anyway this has nothing to do with indiscernibility. *Example 3* In microphysics one assumes that particles of the same kind and in the same state (i.e. with no intrinsic differences) count as equal or equivalent and so can be exchanged both in fact and in the calculations. This assumption is usually cast with the term 'indistinguishable' replacing 'equal' or 'equivalent' – as if the particles could care about our ability to discern among them. The fact is that we do distinguish among them by their extrinsic (spatiotemporal) properties and so are able to count them. See Figure 2.3. In other words, the truth of the matter is that the "elementary particles" are distinct and often distinguishable in practice, but (a) they can be counted as equal or equivalent and (b) when constituting certain wholes they surrender part of their individuality. (See Ch. 5, Sec. 4.2 for the partial loss of "identity" of things when they become components of a system.) In short, Postulate 2.5 is confirmed by microphysics. Plato (*Phaedo* 74a–75c), Leibniz (1704, Ch. XXVII), and Bolzano (1851, Sec. 50) were right in assuming it.

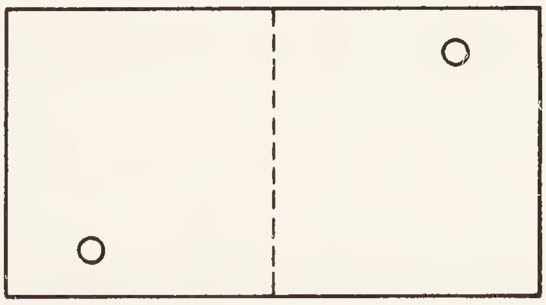


Fig. 2.3. Two electrons in a box: exchangeable if in the same state, yet distinct and distinguishable (discernible). I.e. their exchange has no effect on the states of the system as a whole. But it can happen precisely because the particles are distinct.

Indiscernibility is a special kind of similarity (see Sec. 3.5). Hence, unlike identity, indiscernibility is not transitive. Therefore it does not generate equivalence classes. Instead, indiscernibility generates indiscernibility spaces, also called tolerance spaces (Zeeman, 1962; Schreider, 1975). These spaces have interesting topological properties and occur in the study of perception. We shall take a quick look at them, starting with

DEFINITION A *tolerance space* $\langle T, \sim_s \rangle$ is a set $T \subset S$ of entities together with an indistinguishability (indiscernibility) relation $\sim_s$ on T.

Example Let T be the collection of entities in the visual field of a subject s (in a given state) and let $\sim_s$ be the visual acuity tolerance for the same subject, i.e. (extensionally) the collection of pairs of entities that s cannot distinguish visually from one another. The structure $\langle T, \sim_s \rangle$ is a tolerance space.

The topology of a tolerance space is determined by the collection of neighborhoods defined by

DEFINITION The *vicinity* of an entity $x \in S$ is the collection of entities indistinguishable from x:

$$N(x) = \{ y \in S | y \sim_s x \}$$

Any given set of entities can be covered by such neighborhoods. (Clearly no topology generated in this way is Hausdorff. Hence it has little if any interest in the study of physical space.) What holds for individual entities holds for aggregates of such since these, too, are entities. That is, we could define tolerance spaces for sets of aggregates of entities. But we cannot define indistinguishability between sets of things, except of course conceptual indistinguishability which is the same as identity. Hence most of the results on tolerance spaces obtained by Zeeman (1962) have no application to reality. A case in point is the theorem according to which if $\langle T, \sim_s \rangle$ is an indistinguishability space so is $\langle 2^T, \sim_s \rangle$. In fact we do not perceive or fail to perceive the elements of 2^T, for they are sets.

4. PROPERTIES OF PROPERTIES

4.1. *Identity and Difference of Properties*

How can property identity be characterized, i.e. under what conditions can two properties be said to be the same? We shall begin by examining, and eventually rejecting, two seemingly natural solutions to this problem – only to end up by declaring the problem itself to be ill conceived.

Suppose properties were conceived of in a nominalistic fashion, i.e. as collections of individuals, or of n-tuples of such (recall the Introduction). Then identical properties would be those possessed by exactly the same individuals (cf. Wilson, 1955). That is, every equivalence class $[P]$ of concomitant properties (Definition 2.11) would be just a singleton. But this proposal is not viable, for it obliterates the difference between properties with equal scope, hence between coextensive though not cointensive attributes, such as "fluid" and "viscous", or "living" and "mortal". Hence here as elsewhere nominalism is a failure.

A second possibility that suggests itself is to adopt a paraphrase of Leibniz' law of individual identity (Postulate 2.5), namely this: "Two properties are the same just in case they have the same (second order) properties". Unfortunately we do not know exactly what a second order property might be. We know only what a second order *attribute* or predicate is – namely one that applies to first order predicates and obeys some system of second order predicate logic. Therefore the logical criterion of attribute identity does not help us out.

Our theory of properties concerns only properties of individuals, hence it does not contain a definition, let alone a criterion, of property identity. This is just as well, for nobody needs such a definition and such a criterion. In science, as in everyday life, one uses properties to individuate entities, not the other way around. Therefore one characterizes the identity (or the difference) of things, namely by Postulate 2.5.

For the above reasons the discussions on property identity in the current philosophical literature are wrong headed hence sterile. Moreover they are sloppy because they are not conducted in the context of a full fledged theory of properties. Let us therefore turn to some genuine problems, and in the first place to the problem of the identity of attributes.

The latter problem is genuine and manageable, if only because we manufacture attributes ourselves. Roughly, the criterion of attribute identity is this: "Two attributes are the same iff they are the same (propositional) function and are assigned the same interpretation". Function identity is not enough because one and the same function may be assigned any number of inequivalent interpretations. To obtain a full characterization of an attribute representing a given substantial property we must append the semantic assumption (or "correspondence rule") telling us what property of which entities the given predicate represents.

A more precise characterization of attribute identity is obtained with the help of our theory of meaning (Vol. 2, Ch. 7). In the latter two predicates are the same just in case they have the same sense and the same reference class. If the predicate happens to belong to a definite closed context, such as a scientific theory, then its sense is defined as the union of its principal ideal and its principal filter. And in any case the reference class of a predicate equals the union of the sets occurring in its domain. Therefore two predicates are the same just in case (*a*) the unions of their respective principal ideals and filters coincide, and (*b*) their domains are the same. In sum, we have

CRITERION 2.3 Two attributes (predicates) are the same iff they have the same sense and the same referents.

Needless to say, two predicates are different just in case they are not the same. Moreover their precise difference can be computed with the help of Definition 7.13 in Vol. 2, Ch. 7.

So much for the pseudoproblem of the identity of properties.

4.2. *Property Weight*

Aristotle saw that not all properties are equally weighty and classed them into essences and accidents. He erred of course in reducing essences to the four basic elements, in regarding them as absolute and unchanging, and in believing change (in particular motion) to be just as accidental as location. Galilei discarded the Aristotelian essences for being unknowable and, by investigating the modest but knowable *accidentia* and their lawful interrelations, changed the face of science. (See Shea, 1972, pp. 70–72.)

Although modern science rejects the essence/accident dichotomy it does not brush aside the very distinction between essential and accidental properties. Far from levelling properties it recognizes that some are necessary for others (Sec. 3.3); moreover, it assumes that there are, as it were, degrees of necessity. So much so that, whenever one sets up a theoretical model of a concrete entity, he tries to seize on its salient properties discarding all others as inessential – at least until further notice. Thus the ontological hypothesis that properties have different weights underlies the strategy of model building.

Where theories are scarce but data plentiful, statistical measures of property weight can be used. The most obvious candidate is of course the coefficient of linear correlation. With its help the following measure of relative property weight can be defined. (Cf. Blalock, 1961.) Let P, Q, and R be properties such that R depends (statistically) upon P and Q. Then

$$P \text{ is } \textit{weightier} \text{ than } Q \text{ for } R \text{ iff } |r_{PR}| > |r_{QR}|,$$

where r_{AB} is the linear correlation coefficient between A and B.

As soon as precise functional dependences among variables can be assumed, a more precise and stable measure of relative property weight can be established. Thus if F is a differentiable function of n variables x_i, where $1 \leqslant i \leqslant n$, each of these "independent" variables contributes its own share to F. An obvious measure of the contributions of x_i to F (or of the weight of x_i relative to F) is

$$w(x_i, F) = \int_{D_i} dx_i \left| \frac{\partial F}{\partial x_i} \right|$$

where D_i is the domain of x_i.

Regardless of the precise mode of computation of property weights, they fall under

DEFINITION 2.15 Any function $w: \mathbb{P} \times \mathbb{P} \to [0, 1]$ such that, for all P, Q, $R \in p[T]$, where $T \subset S$,

(i) $w(P, P) = 1$ for all P;

(ii) if P precedes (is necessary for) Q, then $w(P, Q) \geqslant w(Q, P)$;

(iii) if P and Q are incompatible [by Definition 2.4], then $w(P, Q) = = w(Q, P) = 0$;

(iv) if P and Q are concomitant [by Definition 2.11], then $w(P, Q) = = w(Q, P) \neq 0$;

(v) for any given $P \neq Q$, $\sum_{Q \in \mathbb{P}} w(P, Q) = 1$;

(vi) $\ulcorner w(P, R) > w(Q, R) \urcorner$ is interpreted as P being weightier than Q for R,

is called a *property weight*, and its value $w(P, Q) \in [0, 1]$ the *weight* of P relative to Q.

What about essential properties? The traditional answers are:

(i) *nominalism and conventionalism*: there are no essential properties;

(ii) *essentialism*: every entity has some properties which are essential, all others being accidental.

The former doctrine is at odds with the practice of science, where one does not assign all properties the same weight but regards some of them as basic and the others as derived. (For example, some chemical properties are steric, i.e. shape-dependent. And some biological properties depend upon the chemical composition.) As for classical essentialism, it contains a grain of truth – namely the recognition that not all properties are removable tags – but it fails in conceiving of accidents as unimportant and lawless. And the latest version of essentialism, namely modal essentialism (cf. Teller, 1975), is useless because it revolves around the vague notion of necessity elucidated in modal logic – the metaphysical inanity of which will be shown in Ch. 4.

Our own brand of essentialism, which may be called *nomological essentialism*, boils down to the following points:

(i) An essential property is, by definition, one that, far from being stray, participates in some law. (On the other hand an accidental property is a nonessential property.) Now, by Postulate 2.7, all properties participate in some laws. Hence *all properties are essential* – which is to say lawful. In other words there are no accidental properties.

(ii) Although there are no accidental properties, there are *attributes* accidental in a given context. Thus the price of gold is accidental from a chemical point of view, not so from a financial point of view.

(iii) Every law of an entity is a property of it (in the extended sense clarified in Sec. 3.3), hence it is an essential property of the entity.

(iv) The real divide is not the nonexisting distinction between *essentia* and *accidentia* but that between *basic* and *derived* properties. This distinction is inherent in the very notion of property precedence: P is *basic to* Q iff $P \preccurlyeq Q$, or $\mathscr{S}(Q) \subseteq \mathscr{S}(P)$. But if P is basic to Q then P is also basic to the law $\ulcorner \mathscr{S}(Q) \subseteq \mathscr{S}(P) \urcorner$.

Nomological essentialism has the following advantages over its rivals: (*a*) it is consistent with a full fledged theory of properties; (*b*) it fits in with the practice of science; (*c*) it discourages the popular view that the essence of a thing is a part of it, namely its unchanging inner core – which core proves in most cases to be imaginary; and (*d*) it puts an end to the medieval dispute (revived by existentialism) concerning whether essence is prior to existence or conversely.

4.3. *Resultants and Emergents*

Let us return to a consideration of properties in general, such as having red blood cells, rather than individual properties, such as having so many red blood cells at a given moment. Some properties may be called *bulk* or *global* properties because they characterize an entity as a whole. Bulk properties are of two kinds: resultant and emergent (Lewes, 1879). Energy is a resultant or hereditary property in the sense that it is possessed by every part of a thing. On the other hand being stable, being alive, having a certain structure, and undergoing a social revolution, are emergent or nonhereditary properties because they are not possessed by every component of the whole: see Figure 2.4. Accordingly we are justified in proposing

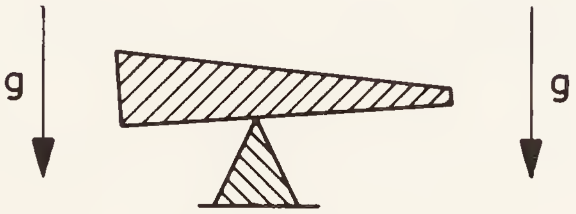


Fig. 2.4. A lever supported by a wedge in equilibrium in a gravitational field. Being in equilibrium is a property of the system as a whole.

DEFINITION 2.16 Let $P \in p(x)$ be a property of an entity $x \in S$ with composition $\mathscr{C}(x) \supset \{x\}$. Then P is a *resultant* or *hereditary property* of x iff P is a property of some components $y \in \mathscr{C}(x)$ of x other than x; otherwise P is an *emergent* or *gestalt property* of x. That is,

(i) P is a *resultant* or *hereditary* property of $x =_{df} P \in p(x) \;\&\; (\exists y)(y \in \mathscr{C}(x) \;\&\; y \neq x \;\&\; P \in p(y))$;

(ii) P is an *emergent* or *gestalt property* of $x =_{df} P \in p(x) \;\&\; (y)\neg(y \in \mathscr{C}(x) \;\&\; y \neq x \;\&\; P \in p(y))$.

Mechanism and individualism assume all properties to be hereditary or resultant, hence explainable by reduction, as is the case with the total charge on a body or the total consumption of goods in a society. We follow instead the tradition of the Greek atomists in recognizing the existence of emergent properties. But, unlike holism, we regard emergents as rooted in the properties of the components, hence as explainable in terms of the latter, though not by reduction to them. For example, temperature is explainable as average kinetic energy of the molecules, but this does not exemplify reduction because averages, though *computed* from individual values alone, are collective properties. Were this not so it would make sense to attribute a temperature value to an individual molecule and a social structure to a person – but it does not. Likewise the cohesiveness of a society can be explained in terms of the participation of its members in a number of social cells but it is societal property not an individual one. These remarks can be generalized into

POSTULATE 2.9 Some of the substantial properties of all composite things are emergent, and all of the emergent properties of a whole are preceded by properties of some of its parts:

(i) $(x)(x \in S \ \& \ \mathscr{C}(x) \supset \{x\} \Rightarrow (\exists P)(P \in p(x) \ \& \ P \text{ is emergent}))$

(ii) $(x)(x \in S \ \& \ \mathscr{C}(x) \supset \{x\} \Rightarrow (P)((P \in p(x) \ \& \ P \text{ is emergent}) \Rightarrow (\exists y)(\exists Q)(y \in \mathscr{C}(x) \ \& \ Q \neq P \ \& \ Q \in p(y) \ \& \ Q \preccurlyeq P)))$.

Whereas the first clause should warm the hearts of holists, the second should cool it. In fact, whereas (i) asserts that some bulk properties are emergent, (ii) suggests that emergents are analyzable – though not reducible. In other words, whereas some bulk properties (the hereditary or resultant ones) are eliminable (definable) in favor of microproperties, others are not – yet in either case they are analyzable or explainable. Let us emphasize that, whereas (ontological) reducibility entails (epistemological) analyzability, the converse is false. There is epistemic novelty in the formation of attributes representing emergent (ontological) novelty. Nor does the explanation of emergence involve the elimination of ontological novelty: a mountain is not explained away when explained as composed of atoms. Explained emergence is still emergence.

4.4. *Properties of Properties*

Every predicate has a number of properties – second order predicates. The n-arity of a predicate is such a second order predicate. Properties such as this one belong in the very definition of the predicate. A second order property of a different kind is that exemplified by the predicate "mechanical predicate". By this one means a predicate, such as the mechanical stress tensor, occurring exclusively in mechanical theories. Such properties are then dependent upon the mode of division of intellectual work. In either case a property of a predicate is a construct not a property of a concrete object.

What about substantial properties: does every such property have properties, and do the latter represent features of a concrete individual? Consider a couple of examples. First: the weight of an organism has the property that it varies in time. But this is just a way of saying that organisms have a *variable weight*, and this is a property of theirs not a second order property. Second: the specificity peculiar to enzymes is (believed to be) a steric property, i.e. one deriving from the shape of the enzyme molecule. But this is just to say that molecule *shape precedes specificity* – and this is a law concerning enzymes, i.e. a first order property.

We make bold and generalize: While every predicate has some (second order) predicates, there is no such thing as a second order substantial property. Whatever we predicate of substantial properties is a predicate – a construct or artifact. (Of course some such predicates are attributed truly and others not, but this is a different story.) This is just as well, for otherwise the symbol '$p(x)$' could not designate the set of *all* properties of x.

5. STATUS OF PROPERTIES

5.1. *The Reality of Properties*

We have tacitly regarded all substantial properties as real, though not as autonomously real, or real in themselves, i.e. apart from the individuals possessing them. More precisely, we have implicitly employed

DEFINITION 2.17 A property P is *real* $=_{df}$ There is at least one individual $x \in S$, other than the null individual, that possesses P (equivalently: $\mathscr{S}(P) \neq \varnothing$).

This definition applies not only to intrinsic properties (represented by unary predicates) but also to mutual properties (represented by n-ary predicates). Thus to say that a certain relation R is real amounts to saying that there are R-related entities or substantial individuals. (Cf. Bolzano, 1837, Sec. 80.5.) We do not assign intrinsic properties a greater ontological weight, or a higher degree of reality, than mutual properties. Thus a mutual property like gravitational interaction (or a certain parent-child mutual action) is just as real as an intrinsic property such as composition (whether chemical or social).

Surely the above view is not the only possible one. Essentially the following doctrines concerning the reality of properties have been put forth:

(ia) *Properties, whether intrinsic or mutual, are real, nay supremely real, and individuals only exemplify them.* (This generalizes Plato's original doctrine of forms, all of which were unary.) This view is untenable, as a property is null unless it is possessed by some individual. Consider in particular a mutual property represented by a binary predicate R. The extension of R is the set of ordered pairs $\langle x, y\rangle$ such that x bears R to y. If the extension is empty, i.e. if R has no examples, then R happens to be the null relation – the one that fails to hold.

(ib) *Whereas intrinsic properties are real, mutual properties are not.* This thesis seems to find support in ancient logic and ancient grammar, with their insistence that all predicates are unary. In turn, this view may spring from a spontaneous tendency to deal with things separately rather than in their mutual relations. It was vehemently defended by Bradley (1893) and still has defenders among those philosophers who regard, say, "South of the Canadian border" as a "spurious property". The thesis has recently been revived in the proposal to split a relation, or arc of a graph, into two demiarcs and a junction or glue – without of course any attempt at characterizing mathematically either of these alleged components (Harary, 1971). Needless to say, the thesis is indefensible and, if adopted, would kill science, which consists largely in an endeavour to disclose connections. And it would render metaphysics quite uninteresting, as it would force us to conceive of reality not as an aggregate of systems but as a mere collection of disconnected individuals. This borders on

(ii) *All properties, whether intrinsic or mutual, are unreal: only individuals are real.* This is of course the nominalist thesis and the exact dual of the generalized Platonist thesis (ia). According to it a property, if intrinsic, is identical with a collection of individuals and, if mutual, is nothing but a collection of ordered n-tuples. This view is muddled: the examples of a relation, i.e. the individual n-tuples, are characterized by their being related. The relata constitute the extension or graph of the relation, not the relation itself. (More against the extensionalist identification of predicates with sets in Vol. 1, Ch. 4, Sec. 1.2, and Vol. 2, Ch. 10, Sec. 1.2.)

(iii) *Neither properties nor individuals are independently real.* Some individuals and some properties constitute things, their states and their changes of state – the only realities. In the fact that *thing b pulls thing c* what is real is neither of the three items separately but the fact as a whole. Only in pure mathematics does one find (or rather create) individuals bereft of all relations, namely the members of structureless sets. And again only in pure mathematics can one maintain that either the individuals or the relations (in particular the functions) constitute the basic objects out of which everything else can be built. Thus set theory takes the view that individuals and collections of such are basic, everything else being reducible to them; and category theory starts on the other hand from the functions which the individuals happen to satisfy. Neither point of view carries over to ontology: here we must regard individuals as well as their properties as so many abstractions. The real thing is the substantial individual with all its intrinsic and mutual properties. Everything else is fiction. Needless to say we champion this third position and do it on the strength of its being inherent in science. Indeed, science has no use for either propertiless individuals or properties that happen not to be properties of substantial individuals.

What holds in general for relations holds in particular for polarities, or relations among polar opposites such as source and sink. A pole is a constituent of a polarity, which in turn is an even degree relation such as a binary interaction between particles differing only in the sign of their electric charge. By definition then poles come in pairs or, generally, in 2 n-tuples: there are no monopoles. (The expression 'electrical monopole', applied to an electrically charged body, is a misnomer.) Therefore it is mistaken to claim that the existence of any one pole renders its partner possible (Stiehler, 1967, pp. 15, 17). However, we shall not set much store by polarity – a pitfall of archaic metaphysics – since most polarities are imaginary: in fact most of them are just not pairs of opposites and others arise from reifying negation. Thus quantity-quality, one-many, subject-object, appearance-reality, and mind-body are not pairs of opposites but of differents. Others, like being-nothingness, identity-difference and the like are genuine pairs of opposites but have no ontic status. Indeed nonbeing does not oppose being except logically, hence they cannot fight except metaphorically. Likewise the opposition between identity and difference is conceptual not a "struggle of opposites". In both cases the confrontation is strictly logical: A "versus" non-A. Which brings us to the important difference between two kinds of relation: binding and nonbinding.

The relation of being greater (or older or richer or worse) than is "external" to its relata in the sense that it does not alter them and does nothing to keep them together or even apart. Similarly with all the spatiotemporal relations such as those of contiguity, betweenness, precedence, and simultaneity: they do not affect their relata. In general, all comparative (order) relations and all equivalence relations are of this kind, i.e. nonbinding. Not so the hydrogen bond or a relation of economic bondage or of cultural influence: these "make a difference" to the individuals so related: they are in a way "internal" to them. This distinction, ignored by Hume and his followers, was pointed out by Peirce (c. 1909, 6.318), who wrote of "existential relations" or relationships in contrast to mere relations. It was also a favorite with Whitehead (1929) and Woodger (1929) who called them 'organic relations' although they also hold among parts of a machine.

Binding relations, i.e. those that "make a difference" to the relata, may be characterized as follows. Two entities $x, y \in S$ are *bound* (or *connected* or *coupled*) iff some changes in x are accompanied (or preceded or followed) by some changes in y. This must be considered a preliminary elucidation, since we do not yet have any exact concept of change. (See Ch. 5, Sec. 4.1.) Nevertheless the previous characterization should suffice to draw the attention to the two main types of substantial mutual property and to avoid some mistakes.

One mistake stemming from overlooking the above distinction concerns the relation between being connected and being a part. Clearly any two things with a common part are connected. But this is a very special kind of connection: overlapping is sufficient but not necessary for connection. Hence, *pace* Leonard and Goodman (1940), and Feibleman and Friend (1945), things need not overlap in order to be coupled. For example, the feudal baron and his serf are connected without having any common part. Another mistake worth warning against is the belief that, since logic does not discriminate between binding relations and nonbinding relations, it ought to be augmented with ontic or physical predicates (Lewis, 1946, pp. 218 ff.). This is not necessary: we may take off the rack any of the n-ary predicates found ready made in logic (and in mathematics) and endow it with the appropriate ontological interpretation. In general it is not advisable to try and supplement logic with ontology: these two disciplines have very different goals and methods, and a single logic should underly all the varieties of ontology.

5.2. *A Critique of Platonism*

The Platonic doctrine of properties is, in a nutshell, the following: (*a*) form is self-existent, ideal, and external to matter, and (*b*) form precedes matter and can take on individuality, or be realized (exemplified) in particulars. Thus a white thing is said to "participate" in the universal Whiteness – or, as logic teachers are still fond of saying, a white individual *instantiates* the predicate Whiteness. (Aristotle corrected this doctrine by denying the autonomy of forms and holding that matter develops form but, to him, the formless substance was real not just a useful fiction as it is for us.) This theory of forms lingers on in a number of contemporary scholars. While only a few (notably Castañeda, 1974) are avowed Platonists most of them would be surprised to learn that they have adopted the least reputable of all of Plato's teachings. Here are a few examples of Platonism picked at random in recent publications.

(*a*) The biologist Paul Weiss (1963) invokes a super-ordinating "principle" hovering above the molecules that make up a cell and regulates it.

(*b*) E. W. Sinnot (1963, p. 194), another biologist, writes that "Form consists of *relations* among particles, or orderly patterns in them. (...) It is a category of being very different from matter, for it is not the nature of the material particles themselves that is involved, but rather how they are related to one another. Form may disappear and appear, as order yields to randomness or comes again, but matter (in its widest sense, as matter-energy) is conservative, moving toward uniformity and maximum entropy".

(*c*) Nobel laureate Werner Heisenberg (1969, pp. 324–325) states: "'In the beginning was Symmetry' – this is surely truer than the Democritean thesis 'In the beginning was the particle'. The elementary particles embody the symmetries, they are the simplest representations of the latter but they are just a consequence of the symmetries".

(*d*) The philosopher U. J. Jensen (1972) has claimed that mental phenomena are not particulars but universals, hence unmeasurable.

A related view is that properties, far from being separate from things, constitute the latter: a thing would be a bundle of qualities – hence a noun could be decomposed into a pack of adjectives. This view, a thoroughly monistic doctrine of form, has been held by the following contemporary thinkers:

(*a*) Russell (1940 p. 98) wrote: "Common sense regards a 'thing' as having qualities, but not as defined by them; it is defined by spatio-temporal position. I wish to suggest that, wherever there is, for common sense, a 'thing' having the quality *C*, we should say, instead, that *C* itself exists in that place, and that the 'thing' is to be replaced by the collection of qualities existing in the place in question. Thus '*C*' becomes a name, not a predicate".

(*b*) A. I. Ujomov (1965, p. 17) states: "The thing is a system of qualities". And the parts of a thing are not parts of space but "parts of a system of qualities" (p. 18).

(*c*) G. Falk (1966, I, p. 77): "A body moving with velocity v is, according to dynamics, nothing but a certain momentum p and a certain amount of energy $\varepsilon = \varepsilon(p)$ that are transported through space with the velocity v". (Disregard the heresy of identifying a function with an arbitrary value of it.)

(*d*) H. Hiż (1971) too has held that any entity can be identified with the set of its properties and moreover that "an individual may be construed as a set of properties of an individual" – pardon the circularity.

Both the Platonic doctrine of forms and the view that things are bundles of qualities are muddled and therefore neither can be formulated exactly. Organic forms are forms *of* organisms; particle and field symmetries are symmetries of properties *of* things (e.g. of their hamiltonians or of their basic laws); a kind of mental phenomena is a class of *individual* mental phenomena; a thing is *characterized* by its properties but not identified with or defined by them, if only because some of its properties are shared by other things; and a set of properties may be used to sketch or model a thing but it does not *replace* the latter. In sum let us heed Ockham's injunction: *Do not detach forms from particulars* – e.g. speak not of Motion but of moving entities. (Or, if you do employ the concept of motion, construe it as the set of moving things.) Likewise let us speak not of Life but of living things, not of Mind but of sensing and thinking things, and so on. Firstly because there is not a shred of empirical evidence for the hypothesis that forms are detachable from their carriers. Secondly because every property is conceptualized as a function defined on some set of entities. (Recall Postulate 2.1 in Sec. 2.1.)

5.3. *The Problem of Universals*

The problem of universals boils down to the questions: What are universals and how do they exist if at all? This problem looks now rather different from what it looked like in the Middle Ages. Indeed we have come to realize that, before rushing to propose an answer such as that universals exist *ante rem* (Platonism), *in re* (Aristotelianism), or *post rem* (nominalism), we should clarify the question itself. To do this we shall define the concept of a universal or rather two radically different concepts of universal in tune with the methodological dualism we adopted in Ch. 1, consisting in splitting every set of objects into entities (S) and constructs (C). Indeed we shall distinguish the *substantial universals*, or widespread properties of entities, from the *conceptual universals* – among them the predicates representing the former. This distinction is introduced explicitly by

DEFINITION 2.18 Let $A \in \mathbb{A}$ be an attribute, $P \in \mathbb{P}$ a property, and T a set. Then

(i) A is a (*conceptual*) *universal* in the set $T \subset C$ of constructs iff the extension of A equals T:

$$A \text{ is } \textit{universal} \text{ in } T =_{df} \mathscr{E}(A) = T,$$

(where the extension function is defined in Vol. 2, Ch. 9 of this *Treatise*);

(ii) P is a (*substantial*) *universal* in the set $T \subset S$ of entities iff the scope of P equals T:

$$P \text{ is } \textit{universal} \text{ in } T =_{df} \mathscr{S}(P) = T.$$

These notions of a universal are then relativized to a set of objects. And since nothing is said about the cardinality of this set, the above definition does not allow one to distinguish a "genuine" universal such as womanhood from an idiosyncrasy such as being the 20th president of México. One might of course try to improve on this situation by introducing a quantitative concept of degree of universality. A suitable candidate for finite sets would be the function characterized by

DEFINITION 2.19 Let $P \in \mathbb{P}$ be a substantial property and $T \in \mathscr{P}(S)$ a nonempty finite set of substantial individuals. Then the *degree of universality* of P in T is the value of the function

$$u\colon \mathbb{P} \times \mathscr{P}(S) \to [0, 1] \text{ such that } u(P, T) = \frac{|\mathscr{S}(P)|}{|T|},$$

where $|A|$ designates the numerosity of A.

This concept would allow one to overcome the primitive dichotomy of the universal and the particular. Still, this more refined concept does by itself not solve the ontological and epistemological problem of universals. To begin with the so called *formal properties of things*, such as number and shape, appear to be full fledged conceptual universals inherent in concrete entities and thus defy the substantial/conceptual dichotomy assumed in Definition 2.18. Thus numerosity is a property of any collection whatever the nature of its components. True, but then a set is a concept. Surely dogs are normally four legged, this being a substantial property of theirs. The corresponding mathematical property is the cardinality of the set of legs of a dog. In other words four leggedness is a substantial property to be distinguished from the mathematical property of the set of legs of a quadruped. The relation between the predicate and the corresponding property is that of representation: the former represents the latter in the case we are considering. Likewise when we say that the history of a person is uninterrupted we do not assign the person a formal property but choose the mathematical property of continuity (a conceptual universal) to represent the smoothness (or absence of gaps) in a sequence of states of an entity. To sum up: not things but our models of them have mathematical properties, and this because we conceptualize substantial properties as functions. This mode of representation is so deeply ingrained in our habits of thought that we often mistake the deputy for his constituency.

In the medieval dispute about universals only conceptual universals seem to have been considered by both the "realists" (Platonists) and the nominalists. In that context we would have sided with Ockham (ca. 1320 I, C.xvi): "nothing is universal except by signification, by being a sign of several things". The conceptual universal "humanity" is not in the individual Socrates but in our idea of him. Of course Socrates is human: he has a property that is universal in the class of humans. This is what the Aristotelians replied: humanness, as a substantial universal, "inheres" in every human being. (In our terminology: the scope of the substantial property of humanness is mankind.) In sum substantial universals (properties and in particular laws) are *in re*. On the other hand conceptual universals (predicates) are *entia rationis*: they are *post rem* if they happen to represent substantial universals preexisting knowledge, and *ante rem* if they anticipate experience or action.

Our theories of substance (Ch. 1) and form (present chapter) allow us to see a further aspect of the problem of universals in a different light, namely the particular/universal dichotomy. First of all this dichotomy, when legitimate, is conceptual not real: the world contains neither bare particulars nor pure forms. Substance and form, individual and universal, are distinct aspects of our conceptual analysis and theoretical modelling of things and facts. Lacking an independent existence, they are not mutually reducible. That is, universals are not collections of particulars (nominalist reduction) and particulars are not bundles of universals ("realist" reduction). The nominalism-"realism" issue, though *de facto* not dead, is finished *de jure*: neither contender was completely right because each focused on just one aspect of the situation while ignoring the other. And each side is done justice by our representing every substantial property by a predicate, and construing the latter as a function from individuals to statements.

Moreover the individual/universal dichotomy, valid and so useful in first order logic, may collapse elsewhere. *Example 1* Sets are neither individuals nor universals – unless they occur as members of a collection of sets, in which case they become individuals, without however the family being a universal. (A nonempty set is constituted by individuals but this does not make it into a universal, much less an individual.) However, some sets are determined or defined by universals (predicates) via the scope function or the principle of abstraction. *Example 2* Some "universals", such as "the set contained in every set", are instantiated by a single object (in this case the empty set). Why call them 'universals' if they are unique? *Example 3* One and the same object can now be regarded as an individual, now as a set (or as a concrete collection). There is nothing final about being an individual. *Example 4* To say that b is an individual is saying too little: what are the properties of b? And saying that P is a property is just as little: what individuals does P apply to, i.e. what is its scope?

The upshot of the preceding discussion is this: Platonism is not a viable metaphysics any more than nominalism is. It might be contended that, though unsuitable as an ontology of substantial individuals, Platonism is the correct philosophy of mathematics because "mathematics has no need for non-classes, like cows or molecules; all mathematical objects and relations can be formulated in terms of classes alone" (Mendelson, 1963, p. 160). This, it is sometimes claimed, is why the von Neumann–Bernays–Gödel version of set theory deals only in classes and sets. Yet (a) the basic predicate of this theory, namely the membership relation, is not construed as a class, and (b) the individuals of the theory are, precisely, sets and classes (as pointed out by Bernays, 1937). Moreover one needs properties even in pure mathematics as soon as one becomes interested in particular sets, i.e. in characterizing them rather than in handling them indiscriminately (i.e. in general) as set theory does. Just think of characterizing, say, the interior of a circle of a given radius without using any predicate. Finally, even if mathematicians one day succeed in producing a mathematics using only classes, metaphysics is likely to continue using the notions of individual and of property if only because both are essential to science: scientific research involves finding out the properties of individuals. Not even the most sophisticated reconstruction of a scientific theory could eliminate properties in favor of classes if only because distinct properties can have the same scope. Nor could it succeed in eliminating individuals, because properties are properties of individuals or of sets of such.

6. CONCLUDING REMARKS

To summarize. There is but one world, denoted □, and it is composed of entities or substantial individuals, i.e. members of the set *S* introduced at the outset (Ch. 1). Every entity is either simple or composite. And every entity has a certain number of properties in addition to its composition. For example a lump of copper, i.e. an entity composed of Cu atoms, is malleable, shiny, melts at 1083 °C, is a good conductor of electricity, is rather scarce, its ore is coveted by all industrial powers, and so on. These are systemic properties of the nonhereditary or emergent kind; on the other hand having a mass and being capable of moving are hereditary properties of a lump of copper, i.e. properties shared by its atomic components.

A substantial property may also be called a *physical* property, in the largest possible sense of 'physical', and in contrast to a conceptual item such as an attribute or predicate. Thus being feathery must be distinguished from the attribute(s) or concept(s) representing featherhood or denying it (e.g. 'featherless'). Firstly because every property is representable by at least one attribute. Secondly because some attributes (in particular the negative and the disjunctive ones) represent no substantial property at all.

Further, there are no properties or forms fluttering over substantial individuals: there are only propertied entities, such as feathery animals. And properties, far from coming either in isolation or in amorphous droves, form clusters and drag one another. That is, every property is either necessary or sufficient for some other properties of the same entity. I.e., there are laws. And these laws are complex properties of entities or, rather, of entire kinds. So much so that they determine the natural kinds. But this is a subject for the next chapter.

CHAPTER 3

THING

So far we have dealt largely with fictions: entities deprived of most of their properties, and forms without a definite stuff. However, we warned that these are fictions and anticipated that they would enable us to construct the notion of a real thing as a fully qualitied individual. This is indeed what a concrete or material object, such as a radio wave or a person or a society is: namely an entity endowed with all its properties, both intrinsic and mutual, permanent and transient.

In the present chapter we shall deepen our study of the concept of a substantial individual and shall investigate in particular the following problems. One of them concerns the characterization of the states of a thing. We shall use the results of this investigation in our treatment of the problem of change in general, to be tackled in Ch. 5. Another question to be dealt with in this chapter is this: What groups different things into natural kinds or species?

Surely the objects we are concerned with "exist in space and time", as the obsolete formula has it. (We shall argue in Ch. 6 that concrete things constitute space and time.) However, we shall still disregard all of the spatiotemporal characteristics of things, characteristics which will be taken up in Ch. 6. Moreover we shall not particularize any properties: that is, our concept of a concrete object, or thing, will apply to a community just as well as to an electron. Finally we shall defer to Vol. 4, Ch. 7 the study of the particular things known as systems. In fact our definition of the system concept will employ the notion of a thing we proceed to elucidate.

1. THING AND MODEL THING

1.1. *Thing*: *Definition*

We stipulate that a thing is an entity or substantial individual (Ch. 1) endowed with all its (substantial) properties (Ch. 2). As may be recalled (Definition 2.3 in Ch. 2, Sec. 3.2) the totality of properties of a substantial individual $x \in S$ was called

$$p(x) = \{P \in \mathbb{P} | x \text{ possesses } P\},$$

where $\mathbb{P}$ is the collection of substantial (nonconceptual) unarized properties.

It may also be remembered that, although we identify or spot any concrete individual by its properties, it is impossible to define an entity as the set of its properties, if only because the expression 'Entity $p(x)$ possesses property P which is a member of $p(x)$' makes no sense. A final caution: although usually a proper subset of $p(x)$ will suffice to *distinguish* x from other entities, nothing short of the totality $p(x)$ of properties of x will *constitute* and *individuate* it, i.e. render it ontically distinct from every other entity. In fact by Postulate 2.5 and its corollary in Ch. 2, Sec. 3.2, what makes a thing what it is, i.e. a distinct individual, is the totality of its properties: different individuals fail to share some of their properties.

The preceding considerations suggest introducing

DEFINITION 3.1 Let $x \in S$ be a substantial individual and call $p(x) \subset \mathbb{P}$ the collection of its (unarized) properties. Then the individual together with its properties is called the *thing* (or *concrete object*) X:

$$X =_{df} \langle x, p(x) \rangle.$$

When referring to a fully qualitied entity, i.e. a thing, we shall often write '$X \in \Theta$' or simply '$x \in \Theta$', where Θ is characterized by

DEFINITION 3.2 The *totality of things* is called Θ:

$$\Theta =_{df} \{\langle x, p(x) \rangle | x \in S \ \& \ p(x) \subset \mathbb{P}\}.$$

This concept of a thing synthesizes the notions of substance and of form studied in Chs. 1 and 2 respectively. We might of course have proceeded the other way around, i.e. starting from things and analyzing them into bare individuals and bunches of properties. But this alternative would have violated the norm that enjoins us to build out of simple units and it would have broken the line of progress we intend to follow.

It might be objected that, since an ordered pair is often defined as a set of sets, namely thus: $\langle a, b \rangle = \{\{a\}, \{a, b\}\}$, our definition of a thing does not retrieve the idea that a thing is a particular. However, it is well known that an ordered pair need not be analyzed and defined as a set but may be characterized as an individual satisfying the single axiom:

$$\langle a, b\rangle = \langle a', b'\rangle \text{ if and only if } a = a' \text{ and } b = b'.$$

Besides, there need be no similitude between a definition and the thing defined. After all, Definition 3.1 characterizes a concept of a thing rather than being the "real definition" of a thing. More on this in Sec. 1.3. And now we turn to some assumptions and a few more definitions.

1.2. *Assumptions*

Our first assumption will be that there are things and indeed infinitely many and moreover a whole continuum of them:

POSTULATE 3.1 The totality Θ of things is an uncountable set.

Remark 1 From a mathematical point of view it would have been more natural to postulate the cardinality of the set S of bare individuals introduced in Ch. 1. That of Θ would have followed because of the one to one correspondence between the two sets. But since bare individuals are fictitious it seemed preferable – from a metaphysical point of view – to assume as little as possible about them. *Remark 2* That an existence hypothesis must be either affirmed or denied explicitly at this point, is shown by the existence of subjective idealism, according to which there are only mental states – or, according to a recent version, just propositions (Fitch, 1971). *Remark 3* Another reason for having to postulate the existence of things is that, if we want to prove anything about existents, we must posit them. We cannot prove the existence of concrete things any more than we can prove the existence of deities or of disembodied minds. What can be proved is that, unless there were things, other items – such as acting on them and investigating them – would be impossible. *Remark 4* That the hypothesis $\ulcorner \Theta \neq \varnothing \urcorner$ is true, is suggested by the fact that all of the factual sciences adopt it tacitly: far from doubting the existence of all things, they take some of them for granted, hypothesize others, and investigate them all. *Remark 5* That the infinity of things is nondenumerable is suggested by the existence of physical continua, such as gravitational fields, all parts of which are fields themselves. Even if the world were to consist of a single particle surrounded by its own gravitational field, it would contain a nondenumerable infinity of things, namely all the parts of the field, which – according to our present knowledge – is infinitely divisible. But of course such a world would contain no detachable parts. *Remark 6* The quantization of a field does not alter the situation: not the field itself but the set of its quanta is denumerable. And even in the absence of quanta a residual or background field remains, namely the one corresponding to the vacuum state. (The latter is not nought.) *Remark 7* The postulated existence of a nondenumerable infinity of things is compatible with the controversial hypothesis that they are organized in a finite number of systems – the only chance for the universe to be spatially finite. But we are not committing ourselves on this point which belongs to Ch. 6 anyhow. In the hypothetical example mentioned in Remark 5 the system is of course the single particle accompanied by its gravitational field.

Having defined the concept of a thing (Definition 3.1) we now define that of the juxtaposition or physical addition of things:

DEFINITION 3.3 Let $X = \langle x, p(x)\rangle$ and $Y = \langle y, p(y)\rangle$ be two things. Then the *juxtaposition* of X and Y is the third thing

$$Z = X \dotplus Y = \langle x \dotplus y, p(x \dotplus y)\rangle.$$

Note that we are not assuming that the properties of the whole are all and only those of the parts, i.e. that $p(x \dotplus y) = p(x) \cup p(y)$. This assumption would be false if only because the compound $x \dotplus y$, for $x \neq y$, has the property of being composed of x and y, which its parts fail to have.

Next we posit that the association theory expounded in Ch. 1, Sec. 1 carries over to things, and we use the symbol '$\dotplus$' to denote the association or juxtaposition of things. More precisely, we lay down

POSTULATE 3.2 Let Θ be the totality of things, $\Box$ a distinguished element of Θ, and $\dotplus$ a binary associative operation in Θ. Then the following holds by stipulation:

(i) $\langle \Theta, \dotplus, \Box\rangle$ is a commutative monoid of idempotents – for any x, y, z in Θ, $x \dotplus y = y \dotplus x$, $x \dotplus (y \dotplus z) = (x \dotplus y) \dotplus z$, $x \dotplus x = x$, and $x \dotplus \Box = \Box \dotplus x = x$;

(ii) $\dotplus$ represents the association or juxtaposition of things;

(iii) the string $x_1 \dotplus x_2 \dotplus \dots \dotplus x_n \in \Theta$, where $n > 1$, represents the concatenate of things x_1 to x_n;

(iv) the neutral element $\square$ of the monoid is the null thing or nonentity – i.e. not a thing proper.

We can retrieve for things everything we defined or proved for bare individuals (or things deprived of their properties other than the property of associating and the properties, such as composition, derived from associability). In particular we can make

DEFINITION 3.4 Let $x, y \in \Theta$ be things. Then x is a *part* of y iff $x \dotplus y = y$. Symbol of the part-whole relation: $\sqsubset$.

Our next assumption is

POSTULATE 3.3 There exists a thing such that every other thing is part of the former. That thing is unique and we call it *the world* or ▯.

Since the world is a thing, and indeed the biggest of all, $x \dotplus \text{▯} = \text{▯} \dotplus x = \text{▯}$.

Because the juxtaposition $x \dotplus y$ includes x and y as parts (by Definition 3.4), the former is the supremum of x and y with respect to $\sqsubset$, i.e., $sup\ \{x, y\} = x \dotplus y$. This remark allows one to prove the analogue for things of Theorem 1.4, namely

THEOREM 3.1 The totality of things has the semilattice structure. More precisely, $\langle \Theta, \dotplus, \square, \text{▯} \rangle$ is a semilattice with least element $\square$ and last element ▯.

The generalization of the notion of concatenation or juxtaposition to an arbitrary set $T \subseteq S$ is that of $sup\ T$. But the existence of this element has got to be assumed:

POSTULATE 3.4 For any subset T of Θ, its supremum $[T]$ exists.

This hypothesis allows us to make

DEFINITION 3.5 The *aggregation* of a set $T \subset \Theta$ of things, or $[T]$ for short, is the supremum of T with respect to the part-whole relation, i.e. the thing such that

(i) $x \sqsubset [T]$ for all $x \in T$;

(ii) if $y \in T$ is an upper bound of T, then $[T]$ precedes y: i.e. if $x \sqsubset y$ for all $x \in T$, then l.u.b. $T \equiv sup\ T = [T] \sqsubset y$.

Remark Just as Postulate 3.2 contained the statement that every finite association of things is a thing (clause (iii)), so the previous definition tells us, among other things, that every association or juxtaposition, whether finite or infinite, yields a thing. This is no ontological triviality: magicians claim to be able to create things out of nothing and to annihilate other things. And some emergentist philosophers believe that certain types of aggregation ensue in objects transcending the realm of things.

The previous assumptions and definitions entail

COROLLARY 3.1 Every part of a thing is a thing:

$$\text{If } x \sqsubset y \text{ and } y \in \Theta, \text{ then } x \in \Theta.$$

Proof The part-whole relation has been defined for things only – not e.g. for concepts – so the assertion that $x \sqsubset y$ is a tacit recognition that both x and y are things.

Remark Nor is this result self evident. Think of the claim that the constituents of ordinary self existents are not things existing autonomously but are the outcome of certain cognitive acts (such as measurements) performed by experimenters with the help of things of a special kind (pieces of apparatus).

Finally we state a consequence of the previous Corollary 3.1 and Postulate 3.3:

COROLLARY 3.2 Every part of the world is a thing:

$$\text{For every } x\text{: If } x \sqsubset \Box \quad \text{then} \quad x \in \Theta.$$

It follows by contraposition that whatever is not a thing fails to be part of the world. But this matter borders on the next subsection.

We close this subsection with two notes of caution. The first is that the world $\Box = [\Theta]$ does not have all the properties of each of its parts: for example, it does not spin, it does not interact with other things, it is not alive, and it has no mind. Indeed, by using Definitions 3.1 to 3.3 we find that

$$\Box = \langle [S], p([S]) \rangle$$

where, in accordance with the warning following Definition 3.3,

$$p([S]) \neq \bigcup_{x \in S} p(x).$$

The second note of caution is that the statement that two or more things form an aggregate satisfying Postulate 3.2, is a factual statement, i.e. one concerning matters of fact. However, this does not render it falsifiable. If any of the clauses of the postulate – e.g. commutativity – should fail, we would conclude not that the postulate has been refuted but that the things of interest have not formed an aggregate. (Equivalently: the resulting thing does not satisfy our axiomatic definition of an aggregate.) For example, if a catalyzer, such as an enzyme, acts on two different reactants, it may happen that it binds faster with one of them than with the other, so that the order in which the three things join does make a difference. But of course this order is temporal, and so has nothing to do with associativity.

1.3. *Thing and Construct*

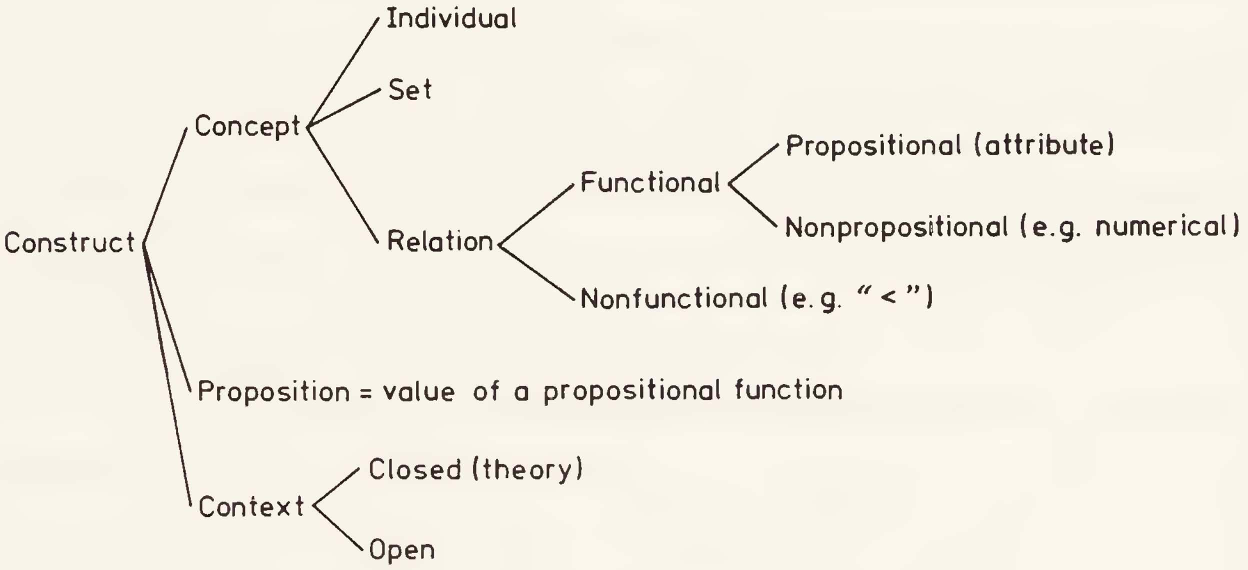
In the previous chapters we have distinguished objects of two kinds: substantial or concrete entities, and constructs. It is now time to sharpen somewhat this distinction – a matter that will be taken up again in Sec. 4.3. To this end we shall start by outlining the concept of a construct, leaving its detailed study to the sciences of constructs, namely logic, semantics, and mathematics.

We feign that there are constructs, i.e. creations of the human mind to be distinguished not only from things (e.g. words) but also from individual brain processes. (Only, we do not assume that constructs exist independently of brain processes.) We distinguish four basic kinds of construct: concepts, propositions, contexts, and theories. Concepts, such as the notion of a thing, are the building blocks of propositions – such as 'All things change' – which are in turn the constituents of contexts – such as the set of all propositions concerning dogs – and of theories.

Basically a concept is either an individual (such as "3" in elementary arithmetic), or a set (such as the set of integers), or a relation (such as "<"). Whether either category is reducible to the other will not concern us here. The most interesting relations are the functions, which are either many-one or one-one relations. We distinguish two kinds of function: propositional and nonpropositional. A propositional function, or predicate, or attribute, is a function that maps some domain of individuals (or n-tuples of such) into some set of propositions. (Recall Ch. 2, Sec. 2.1.) A nonpropositional function takes values in a set other than a set of propositions. (However, every nonpropositional function may be paired off to a propositional function. For example the sine function, which maps reals into reals, is a nonpropositional function that can be assigned the propositional function

$$\begin{aligned}&\text{Sin}: R \times R \to P, \quad \text{where}\\ &\text{Sin}(x, y) = (\sin x = y) \quad \text{for} \quad x, y \in \mathbb{R},\end{aligned}$$

P being the set of all propositions containing the predicate Sin.) Finally, a context is a set of propositions sharing a reference class, and a theory is a context closed under the operation of deduction. In sum we have the following classing:


Individual
Set
Concept
Propositional (attribute)
Functional
Construct
Relation
Nonpropositional (e.g. numerical)
Nonfunctional (e.g. "<")
Proposition = value of a propositional function
Closed (theory)
Context
Open

Constructs do not have all the properties of things. For example sets add and intersect but do not aggregate, do not move around, have no energy and no causal efficacy, etc. Constructs, even those representing things or substantial properties, have a conceptual structure not a material one. In particular, predicates and propositions have semantical properties, such as meaning, which is a nonphysical property. We summarize and extrapolate:

POSTULATE 3.4 Every object is either a thing or a construct, no object is neither and none is both.

Equivalently: Every set O of objects splits into two disjoint sets, one of things, the other of constructs. In the simplest nontrivial case, the set O of objects contains but one thing ϑ and one construct, namely $\{\vartheta\}$. That is, $\Theta = \{\vartheta\}$, $C = \{\{\vartheta\}\}$, $O = \{\vartheta, \{\vartheta\}\}$. Clearly, $\Theta \cap C = \varnothing$.

Substantial individuals devoid of properties, and substantial properties apart from the individual things possessing them, rank as constructs. Likewise a change in properties apart from every changing thing, and a change in an individual apart from a change in its properties, are fictions, i.e. members of some C.

Postulate 3.4 is an axiom of *methodological dualism.* It does not commit us to metaphysical dualism: we are not claiming that there are two kinds of thing, the *res extensa* and the *res cogitans*, or things proper and ideas. We take it that constructs, whether useful or idle, scientific or mythical, are fictions not entities. Hence they are not part of the real world even when they take part in our representations of the latter.

Moreover the preceding postulate entails that each category of objects, Θ and C, has its own peculiarities but does not preclude the possibility that they share some properties. For example some constructs concatenate just like things, and many things are just as thinkable as constructs. However, these two properties are mutual not intrinsic. Intrinsic properties are either of things or of constructs. On the other hand a mutual (e.g. binary) property can link things with constructs. An example of a mutual property of this kind is that of representation, such as it occurs in the statement ⌜Proposition p represents thing b⌝.

The totality of objects has no property other than that of being the union of the class of things and that of constructs. Therefore ontology cannot be construed as the theory of the arbitrary object, or of all objects: there is nothing to be said about O beyond what has just been said. This does not preclude the possibility of making long discourses about the ability to think of all objects, for when so doing we employ binary predicates such as "representing" or "thinking of", which are not subject to the above restriction.

Another obvious consequence of the preceding considerations is that concrete objects (things) have no intrinsic conceptual properties, in particular no mathematical features. This last statement goes against the grain of objective idealism, from Plato through Hegel to Husserl, according to which all objects, in particular material things, have ideal features such as shape and number. What is true is that some of our *ideas* about the world, when detached from their factual reference, can be dealt with by mathematics. (For example, by analysis and abstraction we can extract the constructs "two" and "sphere" from the proposition "That iron sphere is composed of two halves".) In particular, mathematics helps us to study the (mathematical) form of substantial properties. In short, not the world but some of our ideas about the world are mathematical.

Any construct that violates Postulate 3.4 will be declared *metaphysically ill formed.* The attribution of conceptual properties to things, and the attribution of substantial properties to constructs, are in the category of metaphysically ill formed statements. Examples of such metaphysical misfits are ⌜A body is a point set⌝ and ⌜The empty set is grey⌝. Likewise the thesis that a thing is identical with the set of its properties (i.e. $\vartheta = p(\vartheta)$) is metaphysically ill formed because it entails attributing substantial properties to a set, as shown by the following example. Let $p(\vartheta) = \{\text{Drunkard, Cantankerous}, \ldots\}$. Identifying this obnoxious individual with his properties leads to attributing drunkenness, cantankerousness, etc., to the set {Drunkard, Cantankerous, ...}. It could be argued that some metaphysically ill formed statements, such as ⌜Santa Claus has a white beard⌝ and ⌜Number three is perching on a branch⌝ are meaningful in certain contexts. But they do not belong in the body of factual knowledge and it is a matter of indifference whether or not they can be assigned truth values in certain contexts, such as fairy tales – e.g. possible worlds metaphysics.

Let us now turn to a special kind of construct, namely the conceptual sketch or model of a thing.

1.4. *Model Thing*

Theoretical science and ontology handle not concrete things but concepts of such, in particular conceptual schemata sometimes called *model things.* Our construal of a thing as a substantial individual together with the set of all its properties (Definition 3.1) is of course such a model thing – albeit a rather poor one. A richer characterization of a thing is given by a set equipped with specified relations, such as functions or operations. For example, if the thing represented is a force field, then the set will be a portion of a geometrical manifold M – e.g. a region in Euclidean three space – together with a tensor field F on M. Briefly, $\langle M, F\rangle \triangleq$ field. We shall adopt this mode of representation by making

DEFINITION 3.6 Let $X = \langle x, p(x)\rangle$ be a thing of class $T \subseteq \Theta$. A *functional schema* X_m of X is a certain nonempty set M together with a finite sequence $\mathbb{F}$ of nonpropositional functions on M, each of which represents a property of T's. Briefly,

$$X_m =_{df} \langle M, \mathbb{F}\rangle, \text{ where}$$
$$\mathbb{F} = \langle F_i \mid F_i \text{ is a function on } M \;\&\; 1 \leqslant i \leqslant n < \infty\rangle.$$

The base set M will be denumerable or nondenumerable, as the case may be. It may or may not be thought of as mapped on a subset of physical spacetime. (In systems theory M is usually taken to be a set of time instants.) As for the functions F_i in $\mathbb{F}$, every one of them refers to the individual $x \in T$ of interest and is thus evaluated at a fixed x in T even though it may be a function of whatever other variables may be the case. The finiteness of the set of components of $\mathbb{F}$ agrees with the part of Postulate 2.3 specifying that there are finitely many general properties (such as length or longevity). And it does not contradict the second part of that postulate, according to which $p(x)$ is nondenumerable for each $x \in S$. Indeed, recall that a single continuous general property, such as age, gives rise to infinitely many individual properties (such as successive ages) as the property in question takes its values. It might be objected that restricting the number of components of $\mathbb{F}$ is artificial since any function, even the constant function, can be decomposed into infinitely many functions, as is the case with Fourier series expansions. But such mathematical tricks have no ontological significance: they are purely conceptual.

Example 1 The simplest functional model of a corpuscle with variable mass is the classical mass point. Here $M = F \times T$, where F is the set of reference frames and $T = \mathbb{R}$ the real line, every point of which is interpreted as an instant of time. And $\mathbb{F} = \langle \mu, \pi, \varphi \rangle$ is a triple of functions on $M = F \times T$, such that $\mu(f, t)$ represents the mass, $\pi(f, t)$ the position, and $\varphi(f, t)$ the force acting on the corpuscle, relative to frame $f \in F$, at time $t \in T$. This functional schema of a particle is consistent with a number of alternative equations of motion (law statements), i.e. it can be shared by a number of different theories hinging on the same model object. *Example 2* A simple electrical model of the brain consists of a finite region of a four manifold (namely spacetime) together with a single function on it representing the electric potential. And a possible chemical model of the same thing is constituted by the same set together with n functions with domain on it, each of which represents the concentration of a given chemical at a given point in space and time.

Finally we assume that the set of functional schemata is nonempty and can grow indefinitely. That is, we lay down the following methodological

POSTULATE 3.4 Any thing can be modelled as a functional schema: For any $X \in \Theta$ there is at least one $X_m = \langle M, \mathbb{F} \rangle \in C$ such that $X_m \triangleq X$.

Remark 1 The failure to distinguish the thing represented from its model is not just a form of mental derangement: it is also at the root of black magic and subjectivism. The idealist who does not distinguish a thing from any of its models cannot account for the multiplicity of schemata of one and the same thing. Consequently he cannot understand the history of theoretical science, which consists partly in the replacement of some schemata by others. *Remark 2* Our characterization of a functional schema is consistent with current scientific practice. However, two notes of caution are in order. Firstly, more sophisticated modes of representation are conceivable, e.g. with the help of the concept of a category (cf. Padulo and Arbib, 1974). Secondly, sometimes a thing is *defined* as a certain relational structure, i.e. it is identified with one of its model objects. This is mistaken, for only (some) constructs can be defined: things can only be represented (or misrepresented), and occasionally also manipulated. But the mistake is harmless if we are aware of it. *Remark 3* Nominalists are bound to dislike our construal of a sketch or schema of an individual thing as a structured set: they may demand "individualism" for both the world and our conceptual representations of it. Thus Goodman (1956, p. 16): "Nominalism for me consists specifically in the refusal to recognize classes"; it "requires only that whatever is admitted as an entity at all be construed as an individual" (ibid., p. 17). (In particular it does not require the concrete/conceptual distinction (ibid., p. 16).) That demand is not met by science, so we can ignore it. *Remark 4* The set M, if infinite, may represent an actual infinity such as a gravitational field or some other continuum. Such an infinity will be looked at askance by Aristotelians and empiricists alike. On the other hand it will warm the hearts of Leibnizians. Thus Bolzano's *Paradoxien des Unendlichen* (1851) carried as a motto the following quotation from Leibniz: "Je suis tellement pour l'infini actuel, qu'au lieu d'admettre que la nature l'abhorre, comme l'on dit vulgairement, je tiens qu'elle l'affecte partout, pour mieux marquer les perfections de son auteur".

The various functional schemata of a given thing need not be equivalent: for example they may exhibit different amounts of structure. Thus a set whose elements are related pairwise is more structured, integrated, or tightly knit, than one whose elements have only unary properties. Moreover an amorphous set does not become structured by increasing the number of properties of its elements even though it does become more complex. (I.e. the more structured the more complex but not conversely.) This suggests measuring the amount of structure of a relational structure by the number and rank of the predicates included in it. A translation of this idea into our terminology is as follows:

DEFINITION 3.7 Let $X_m = \langle M, \mathbb{F} \rangle$ be a functional schema of a thing X and call the *rank* of $F_i \in \mathbb{F}$ the number of independent variables or argument places of F_i. Then the *amount of structure* of thing X exhibited by its functional schema X_m equals

$$\alpha(X_m) = \sum_k n_k (r_k - 1),$$

where n_k is the number of functions of rank r_k in $\mathbb{F}$.

Now, two relational structures, such as $X_m = \langle M, \mathbb{F} \rangle$ and $X'_m = \langle M', \mathbb{F}' \rangle$, are said to be *similar* iff $\mathbb{F}$ and $\mathbb{F}'$ have the same number of components and their corresponding components are of the same rank (Tarski 1954). In our terminology: two model things X_m and X'_m are *similar* iff their amounts of structure are the same, i.e. if $\alpha(X_m) = \alpha(X'_m)$. This notion suggests attributing the modeled things themselves the property of being similar:

DEFINITION 3.8 Let X and X' be two things modeled as $X_m = \langle M, \mathbb{F} \rangle$ and $X'_m = \langle M', \mathbb{F}' \rangle$ respectively. Then X and X' are *formally* (morphologically) *similar under the same modelings* iff their respective functional schemata X_m and X'_m are similar (i.e. if they exhibit the same amount of structure).

Remark 1 In particular, things described with the help of the same mathematical formalism are modeled similarly and can thus be declared to be similar – under the given modeling! (See Bunge, 1973a, Ch. 5.) However, since morphological similarity as defined above is not an intrinsic property of things but rather a property of pairs thing-model thing, we must be careful not to mistake the above concept for that of objective similarity introduced in Ch. 2, Sec. 3.5. Whereas the latter is an ontological concept the former is an ontological-epistemological one. *Remark 2* It goes without saying that things belonging to the same natural kinds – e.g. electrons, neurons, peasant societies – are represented by the same functional schemata. In particular, indiscernibles are representable by the same model things. In other words, for theoretical purposes we may treat indiscernibles *as if* they were identical – which of course they are not (cf. Ch. 2, Sec. 3.6).

So much for a preliminary characterization of model things, or sketches of things. We proceed to refine the notion of a functional schema of a thing, by taking a closer look at the n-component function $\mathbb{F}$. This will enable us to elucidate the notion of a state of a thing, a notion which will be crucial to our investigation of change in general.

2. STATE

2.1. *Centrality of the State Concept*

Every thing is – at a given time associated with a given reference frame – in some state or other. This is a hypothesis about the furniture of the world and, since it does not specify either the kind of thing or the kind of state, it is ontological. It is not a self evident assumption. For one thing it is rarely if ever formulated explicitly. For another the world might be constituted in such a way that the assumption be false. Moreover according to the classicist interpretations of the quantum theory the latter fails to assign definite states to its referents. But this is mistaken: the truth is that the theory does not assign its referents classical states such as sharp position and velocity values, but quantum states such as position and velocity probability distributions.

The ontological hypothesis that every thing is in some state or other underlies all science, has invaded philosophy, and has spilled over to ordinary knowledge, to the point that statesmen speak of the state of the nation. Each exact scientific theory is of course in a position to give a precise characterization of the concept of a state. But it will be concerned with a particular kind of state, such as the dynamical or the chemical or the physiological or the economic state of its referents, not with the generic concept of state of a thing. To elucidate the latter concept is the business of exact ontology, in particular of systems theory.

Surprisingly enough, systems theory has so far failed to give an exact and general enough analysis of the concept of the state of a system. To begin with the usual accounts (e.g. Zadeh and Desoer, 1963; Mesarović and Takahara, 1975; Padulo and Arbib, 1974) do not apply either to continuous systems such as fields or to quantum-mechanical systems. Moreover the concept of a reference frame has no place in those accounts, probably because it is not needed in automata theory, electrical network theory, and a few other theories. Yet the concept is central to many other theories. (Recall Example 1 in Sec. 1.4.) For example, it occurs in the explanation of the working of an electric motor – not to speak of the underlying theories of mechanics and electrodynamics. The notion of a reference frame is in fact so central in physics that the states of any real physical system are relative to some frame or other: just think of the state of motion of a body, or of the state of an electromagnetic field.

A second concept one misses in systems theory is that of a law. To be sure a general theory of things should neither presuppose nor contain any specific (e.g. thermodynamical or chemical or biological) law, for otherwise it would not be so general. (Equivalently, the systems it describes could not be realized in a number of alternative ways, in particular employing materials of different kinds – hence satisfying different laws.) However, the general concept of a law – the philosophic concept as distinct from a particular law statement such as Ohm's – ought to play some role in systems theory if only for a mathematical reason. Indeed, but for the laws – which place restrictions upon the ranges of the components of the function $\mathbb{F}$ occurring in any functional schema of a thing – we should accept the usual characterization of the collections of possible states (or state spaces) as vector spaces or even as inner product or metric spaces. The existence of laws ruins this characterization. In fact if a law restricts the range of a variable, as it usually does, then it is no longer true that the product of a variable by an arbitrary scalar belongs to the same space. In sum, not even the most general scientific theories – namely those that qualify also as metaphysical theories – have given satisfactory characterizations of the general concept of a state.

Nor have metaphysicians exactified the notion of state – nor, a fortiori, that of change of state. Not even the process metaphysics – such as those of Hegel, Bergson, Whitehead, and their numerous followers – have taken it upon themselves to clarify the notions of state and of change of state – although in science, at least, a process is a sequence of different states. Consequently those philosophers have bequeathed us obscure philosophies of change. Much the same holds for the fashionable possible worlds metaphysics: here the concept of state of a possible world is central, yet it remains undefined. Likewise the various systems of inductive logic employing the concepts of state and of state description fail to analyze them in a manner consistent with the concepts of state occurring in science.

Such a lack of analyses is deplorable if only because the concept of a state, hence the notions of a state description and of a change of state, pose interesting philosophical problems. For example, since every state of a thing is determined by a bunch of state variables for the thing, and since the choice of the latter is partly determined by the state of the art, what right do we have in assigning a thing an objective state portrayed by a state description? Another example: the so-called identity theory (or rather hypothesis) is usually formulated as stating the identity of mind and brain. Would the hypothesis not gain in precision by construing it as stating that mental states are states of the nervous system (or rather of subsystems thereof) and therefore mental events changes in the states of neuron assemblies? These and other philosophical problems call for, even before they are properly stated, an elucidation of the notion of a state of a thing. Let us therefore turn to this task.

2.2. *State Function*

Recall the notion of a functional schema $X_m = \langle M, \mathbb{F} \rangle$ representing a thing X (Definition 3.6). It consists of a set M and a list $\mathbb{F} = \langle F_i | 1 \leq i \leq n \rangle$ of n functions with unspecified domain M and equally unspecified codomains V_i. Each component $F_i : M \to V_i$ of $\mathbb{F}$ is supposed to represent a general property, or property of things of the kind to which X belongs. And each value of F_i for a particular entity at a point $x \in M$ is supposed to represent an individual property, or property of X. In Example 1 of Sec. 1.4 the function $\mu : F \times T \to \mathbb{R}^+$ represents the mass (a general property), and its value $\mu(f, t)$ represents the mass of the given corpuscle, relative to frame f, at time t. Unlike the F_i, which are property-representing "variables" (or rather functions), f and t are not "variables". They are arbitrary members of certain sets (F and T respectively) and they are not possessed by any thing in particular: on the contrary they are rather public in the sense that they can be "used" by a number of things.

The components F_i of the list $\mathbb{F}$ of functions in a functional schema are usually called *state variables* because their values contribute to characterizing or identifying the states the thing of interest can be in. We shall call them *state functions* because this is what they are. And we offer the following preliminary characterization: A function is a *state function* for a thing of a given kind if it represents a property possessed by the thing. Whether this representation is faithful (true) is immaterial to the function's qualifying as a state function. What is decisive is that the function refers to the thing and can be interpreted as representing or conceptualizing the intended property.

A few more examples should give a better feel for the concept of a state function.

Example 1 Qualitative global property: social structure. Let $F_i \triangleq$ social structure in ith respect (occupation, or income, or number of school years, or political power, or whatever). Call Σ the set of all social systems (communities or organizations) and $T \subseteq \mathbb{R}$ the set of instants. Further, call $\mathscr{C}(\sigma)$ the (atomic) composition of system $\sigma \in \Sigma$ at the person level, i.e. the set of individuals composing σ. Define on $\mathscr{C}(\sigma)$ a social equivalence relation $\sim_i$ as that which induces the partition of $\mathscr{C}(\sigma)$ into groups of people that are homogeneous in the ith respect (roughly the same occupation, or the same income, etc.). Call $\mathscr{C}(\sigma)/\sim_i$ such a partition. Then a social structure is a list $\mathbb{F}$ of n state variables of the form

$$F_i: \Sigma \times T \to \{\mathscr{C}(\sigma)/\sim_i | \sigma \in \Sigma\}.$$

Example 2 Population, a conspicuous biological and sociological variable, may be conceptualized as a state function

$$F: \Sigma \times T \to \mathbb{N}$$

from pairs $\langle$community σ of organisms of a kind, instant $t\rangle$ to the natural numbers $\mathbb{N}$. Each value $F(\sigma, t) = n$, for $n \in \mathbb{N}$, represents an individual property of σ, whereas the function F itself represents a general property, or a property of all the members of the set Σ.

Example 3 Quantitative stochastic property. Let $F \triangleq$ Momentum probability distribution. Here $M = \{q\} \times \{f\} \times \mathbb{R}^3 \times T$ and $V = \mathbb{R}$, where $q \in Q$ is a quantum-mechanical entity of kind Q, $f \in F$ a reference frame of kind F, and $\mathbb{R}$ the real line. The function $F: M \to V$ satisfies a certain law statement and '$F(q, f, p, t)\, dp$' is interpreted as "The probability that entity q, relative to frame f, at time t, possesses a momentum comprised between p and $p + dp$". (In rather common notation, $F = |\varphi|^2$, where φ is the Fourier transform of the state function ψ representing the position distribution.)

We are now ready for a general characterization of our concept:

DEFINITION 3.9. Let X be a thing modeled by a functional schema $X_m = \langle M, \mathbb{F}\rangle$, and suppose that each component of the function

$$\mathbb{F} = \langle F_1, F_2, \ldots, F_n\rangle: M \to V_1 \times V_2 \times \ldots \times V_n$$

represents a property of X. Then F_i, for $1 \leqslant i \leqslant n$, is called the *ith state function* for X, $\mathbb{F}$ the *total state function* for X, and its value

$$\mathbb{F}(m) = \langle F_1, F_2, \ldots, F_n \rangle (m) = \langle F_1(m), F_2(m), \ldots, F_n(m) \rangle$$

for $m \in M$ is said to represent the *state of X at m* – all this *in the representation X_m*. Furthermore, if all the V_i for $1 \leqslant i \leqslant n$ happen to be vector spaces, then $\mathbb{F}$ is called a *state vector* for X in the representation X_m.

Notice the cautious expression 'in the representation X_m'. The reason is that there is no such thing as the absolute state function for a given thing: indeed there are as many state functions as functional schemata of the thing can be conceived, i.e. any number of them. (For example, whereas lagrangian theories employ generalized coordinates and velocities as the basic state functions, hamiltonian theories choose generalized coordinates and momenta.) And even one and the same representation is compatible with infinitely many choices of systems of units, every one of which may ensue in a different state function.

The sole test for the adequacy of a choice of state functions is the adequacy (factual truth) of the theory as a whole, in particular that of its key formulas, which are those interrelating the various components of the total state function – namely the law statements and constraint formulas of the theory. Even so there may be alternative though basically equivalent formulations of one and the same theory. (For example, most field theories can be formulated using either field strengths or field potentials, and the latter are mutually equivalent modulo certain arbitrary constants or even functions.) In the case of equivalent theories there are no preference criteria other than those of computational convenience, or else heuristic power, or even sheer beauty or fashion.

To put it negatively: The choice of state functions is not uniquely determined by empirical data but depends partly on our available knowledge, as well as upon our abilities, goals, and even inclinations. This consideration will play an important role in any talk of states and state spaces, of which more in Sec. 2.4. But lest we have given the impression that the choice of state functions is totally arbitrary and a matter of sheer taste, let use hasten to add that whatever set of state functions be chosen, they are supposed to abide by certain law statements – and this is not a matter of convention. More on this in the next section.

Finally a point of mathematical detail. Typically the state functions for a given thing are defined on a single domain M. Think of the mass density, stress tensor, velocity field, entropy density, and the other "variables" concerning a fluid. Likewise the set of dynamical variables ("observables") of a quantum mechanical system are all operators on the Hilbert space associated with the system. However, it might happen that a given set of state functions for things of a given kind fail to be defined on the same domain. In this case a harmless trick will allow us to assign them all the same domain and thus comply with a tacit condition of Definition 3.9. Thus if

$$F_1: A \to B \quad \text{and} \quad F_2: C \to D, \quad \text{where} \quad A \neq C \quad \text{and} \quad B \neq D,$$

we can adopt the new state functions

$$\left.\begin{array}{lll} F_1': A \times C \to V_1 & \text{such that} & F_1'(a, c) = F_1(a) \\ F_2': A \times C \to V_2 & \text{such that} & F_2'(a, c) = F_2(c) \end{array}\right\} \quad \text{for all} \quad a \in A, c \in C.$$

But, as we noted before, it is unlikely that we shall ever have to make use of this artifice.

2.3. *Law Statements as Restrictions on State Functions*

In Ch. 2, Sec. 3.3, we gave an exact but not very illuminating definition of a law. Roughly, we stated that if P and Q are two properties of entities of a kind, then any inclusion relation between the scope of P and Q counts as a law. We are now in a position to offer a deeper analysis of the law concept, namely as a condition on certain state functions for some thing. Such conditions may take a number of forms, depending not only on the things themselves but also on the state of our knowledge. All of the following are conspicuous simple forms:

Example 1 Range of $F = V$, where F is a state function and V some well defined set. (Think of the relativistic restriction on the velocity values.)

Example 2 $\partial F/\partial t \geqslant 0$, where $t \in T$, and $T \subseteq \mathbb{R}$ occurs in the domain of F.

Example 3 Kinematical model of a thing (usually called dynamical system):

$$\frac{d\mathbb{F}}{dt} = \mathbb{G}(\mathbb{F}, t), \text{ with } \mathbb{G} \text{ a specific function.}$$

Example 4 Lagrangian model of a system:

$$\int_{t_1}^{t_2} dt\, F(q, \dot{q}, t) = \text{extremal, with } t_1,\ t_2 \text{ two selected elements of } T \subseteq \mathbb{R}.$$

Example 5 Field theoretic model of a system:

$$F_2(x, y) = \int_V du\, dv\, F_1(u-x, v-y), \text{ with } F_1, F_2\colon E^3 \times E^3 \to \mathbb{C}.$$

Example 6 Another model of the same genus:

$$\nabla^2 F_1 = F_2, \quad \text{with} \quad F_1, F_2\colon E^3 \to \mathbb{R}.$$

The preceding considerations suggest adopting the following characterization:

DEFINITION 3.10 Let $X_m = \langle M, \mathbb{F} \rangle$ be a functional schema for a thing X. Any restriction on the possible values of the components of $\mathbb{F}$ and any relation among two or more such components is called a *law statement* iff (i) it belongs in a consistent theory about the X's and (ii) it has been confirmed empirically to a satisfactory degree.

(It might be objected that, since we know whether a function is a state function only once we have satisfied ourselves that it occurs in some law statement, the above definition is circular. Not so, because we have not *defined* state functions in terms of law statements. All we have asserted is that lawfulness is a test or criterion of state functionality.)

If a law statement concerns a certain thing x, we call it $L(x)$, and we may call $\mathbb{L}(x) = p(x) \cap \mathbb{L}$ the totality of laws, or rather law statements, for x. Likewise we call $L(T)$ a law possessed by every thing in the set T, and $\mathbb{L}(T) = p(T) \cap \mathbb{L}$ the totality of laws "obeyed" by everything in T. This is not a mere matter of notation for, as we saw in Ch. 2, Sec. 3.3, laws are properties. And, being properties, they are representable as functions. Indeed, a law statement may be construed as the value of a certain function – a *law function* – with domain the class T of things concerned, and codomain the set of law statements of the form $L(x)$. In short,

$$\text{If } T \subset \Theta, \quad \text{then} \quad L\colon T \to L(T).$$

For example, Ohm's law for a battery-resistor circuit x (an individual of a certain class C) can be written

$$L: C \to L(C), \quad \text{where} \quad L(x) = \ulcorner e(x) = R(x) \cdot i(x) \urcorner$$
$$\text{for every} \quad x \in C,$$

where in turn e, R and i are real valued functions on C representing the electromotive force, the resistance, and the current intensity respectively. This manner of writing shows clearly that the laws are properties of things and some of them interrelate properties of things.

The same ideas can be represented in an alternative way that lends itself more easily to generalization, namely as follows. Consider again the set C of battery *cum* resistor circuits. As a state vector for things of this kind we may choose either the voltage-current pair or the voltage-charge pair. We pick the former:

$$\mathbb{F} = \langle e, i \rangle : C \to \mathbb{R} \times \mathbb{R} \quad \text{where } \mathbb{R} \text{ is the set of real numbers.}$$

This function assigns to each thing $x \in C$ the pair of real values $\langle e(x), i(x) \rangle$, in such a way that $e(x) = R(x) \cdot i(x)$, where $R(x)$ is a real number. We define now the first and second projections, G and H, of $\mathbb{F}$, as follows:

$$G: \mathbb{R} \times \mathbb{R} \to \mathbb{R}, \qquad H: \mathbb{R} \times \mathbb{R} \to \mathbb{R}$$

such that

$$\langle e(x), i(x) \rangle \overset{G}{\mapsto} e(x)$$

$$\langle e(x), i(x) \rangle \overset{H}{\mapsto} R(x) \cdot i(x)$$

$$\begin{tikzcd}
 & C \arrow[dl, "G \circ \mathbb{F}"'] \arrow[d, "\mathbb{F}"] \arrow[dr, "H \circ \mathbb{F}"] & \\
\mathbb{R} & \mathbb{R} \times \mathbb{R} \arrow[l, "G"] \arrow[r, "H"'] & \mathbb{R}
\end{tikzcd}$$

Since

$$(G \circ \mathbb{F})(x) = e(x) \quad \text{and} \quad (H \circ \mathbb{F})(x) = R(x) \cdot i(x),$$

Ohm's law now reads

$$G \circ \mathbb{F} = H \circ \mathbb{F},$$

i.e. as a statement of the identity of two function compositions.

The previous treatment carries over to any other situation where a fixed thing is being considered and the domain of the state function $\mathbb{F}$ is taken to be an arbitrary set M, such as the collection of all instants, or a portion of a certain manifold. We could multiply the examples, analyzing more and more complex cases, but this would not lead us very far. It is far more profitable to look into general frameworks for scientific theories, such as lagrangian and hamiltonian dynamics (for which see Bunge, 1967b) – but we must get on with the matter of states.

2.4. *State Space*: *Preliminaries*

Every theoretical model of a thing is concerned with representing the really possible (i.e. lawful) states, and perhaps also the really possible (lawful) changes of state, of the thing. A few typical examples should give us a feel for this matter.

Example 1 Let N be a set of neurons, or of nerve fibres, and suppose that each such unit can be in either of two states: *on* (firing) or *off* (not firing). That is, we can introduce a state function $F: N \to \{0, 1\}$ representing the neuronal activity, such that

$$\text{For each } n \in N, F(n) = \begin{cases} 1 & \text{iff } n \text{ is on} \\ 0 & \text{iff } n \text{ is off.} \end{cases}$$

The functional schema of the neural system is then $\langle N, F\rangle$ with N and F as described above. And the corresponding state space for each neuron $n \in N$ is $S(n) = \{0, 1\}$. If the neurons are regarded as mutually independent, the state space of the aggregate $[N]$ of neurons (i.e. the system of interest) is $S([N]) = \{0, 1\}^{|N|}$. Thus for a system composed of 3 neurons, the state space has 2^3 elements: $S = \{\langle 0, 0, 0\rangle, \langle 1, 0, 0\rangle, \langle 0, 1, 0\rangle, \langle 0, 0, 1\rangle, \langle 1, 1, 0\rangle, \langle 1, 0, 1\rangle, \langle 0, 1, 1\rangle, \langle 1, 1, 1\rangle\}$. *Example 2* In the genetics of populations, three state functions are often employed: the size N of a population, the probability p of some particular gene, and the latter's adaptive value v. Hence for a system composed of two interacting populations, A and B, the state space is the region of $\mathbb{R}^6$ spanned by $\langle N_A(t), N_B(t), p_A(t), p_B(t), v_A(t), v_B(t)\rangle$ in the course of time. *Example 3* In the elementary theory of the ideal gas, the state function is the triple pressure-volume-temperature functions. The corresponding state space is a cube contained in $(\mathbb{R}^+)^3$. *Example 4* In chemical kinetics the instantaneous state of a chemical system is described by the values of the partial concentrations of both reactants and products. Therefore the state space of the system is inside $(\mathbb{R}^+)^n$, where n is the number of system components (reactants, catalyzers, and products). *Example 5* In elementary electrostatics the state function is $\mathbb{F} = \langle \rho, \varphi\rangle$, where ρ represents the electric density and φ is the electric potential. Hence the local state of the given field is the value of $\mathbb{F}$ at $x \in E^3$, where E^3 is Euclidean three space. And the entire state space is the set of ordered couples $\{\langle \rho(x), \varphi(x)\rangle \in \mathbb{R}^2 | x \in V \subseteq E^3\}$, where V is the spatial region occupied by the field. *Example 6* In quantum mechanics the state of a system is represented by a one dimensional subspace (or ray) of the Hilbert space assigned to the system. Since a thing of this kind is typically not attributed a pointlike location but assumed instead to be spread over some spatial region $V \subseteq E^3$ (with a definite probability distribution), the state of the thing is the set of all values its state vector takes in V.

Before attempting to jump to any generalizations let us emphasize a point of method made in Sec. 2.2. We may certainly assume that, whether we know it or not, each thing – in particular each isolated thing – is in a definite state relative to some reference frame and at each instant (or else point in spacetime). Yet our *representation* of such a state will depend upon the state function chosen to represent the thing – which choice depends in turn upon the state of our knowledge as well as upon our goals. What holds for each single state holds a fortiori for the entire state space for a thing. That is, far from being something out there, like physical space, a state space for a thing stands with one leg on the thing, another on a reference frame, and a third on the theoretician (or modeller). To persuade oneself that this is so it is enough to take a new look at Example 5 above, where a reference frame at rest relative to the field source was assumed. If the system is now considered relative to a moving frame – moving, that is, with respect to the field source – a four vector current density will have to replace the single charge density, and a four vector potential will take the place of the single scalar potential. (Alternatively – and this is where the scientist's freedom comes in – the four potential can be replaced by an antisymmetric tensor representing the electric and magnetic components of the field relative to the chosen frame.)

Having emphasized the conventional ingredient of every representation of states, let us now stress that it has an objective basis as well. This can be seen from the existence of two different concepts of state space, one of which is more realistic than the other. Suppose we have settled on a certain total state function $\mathbb{F}$. If we form the cartesian product of the codomains of the various components of $\mathbb{F}$ (in tune with Definition 3.8), we obtain the codomain V of $\mathbb{F}$ itself, a set that will be called the *conceivable state space* for the thing represented. This is precisely what we did in the examples that sparked off this section.

However, a state function may not take values in its entire codomain but may be restricted to a subset of the latter, and this by virtue of some law. (Recall Sec. 2.3.) Therefore for every component of $\mathbb{F}$ our concern should be the range of it rather than its codomain. For example, the total population of organisms of a given kind in a given territory is constrained not only by the carrying capacity of the latter but also by the birth and death rates, as well as by additional factors such as sunshine and rainfall. Again, although the range of the speed function for a body is the entire real interval $[0, c)$, where c is the speed of light in a void, an electron traveling in a transparent medium will not come close to the upper bound c, for it is subject to further laws. In general: Only those values of the components of the total state function that are compatible with the laws will be really (not just conceptually) possible. In other words, because the laws impose restrictions upon the state functions and their values, hence upon the state spaces, only certain subsets of the latter are accessible to the thing represented. We shall call the accessible part of the state space the *lawful state space* of the thing in the given representation and relative to a given frame. (It is sometimes called the space of physical states: cf. Hirsch and Smale, 1974.) To say that a thing *behaves lawfully* amounts then to saying that the point representing its (instantaneous) state does not wander beyond the bounds of the state space chosen for the thing (See Figure 3.1).

2.5. *Definition of a State Space*

The preceding remarks can be summarized in terms of the concepts introduced in Sec. 2.3:

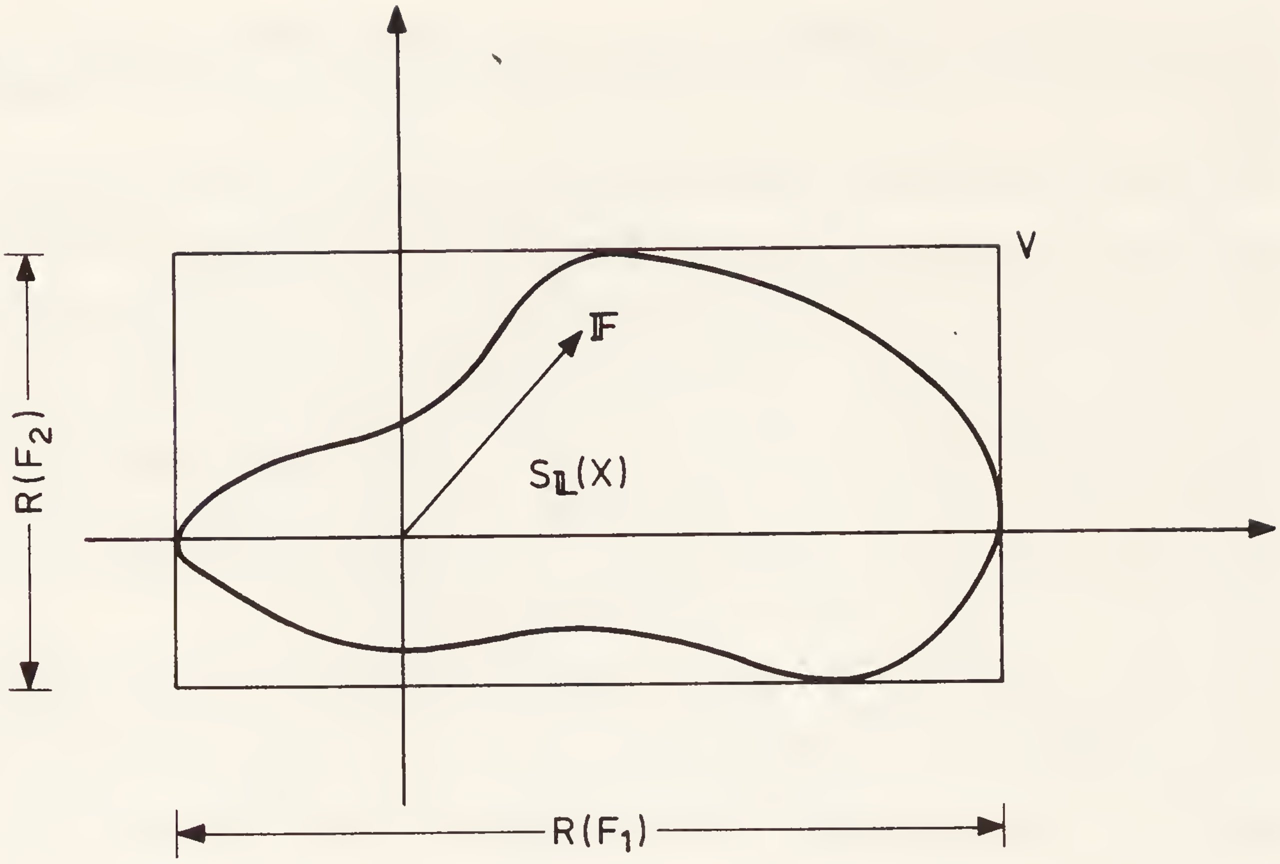


Fig. 3.1. The lawful state space $S_{\mathbb{L}}(X)$ for a thing X is a subset of the cartesian product of the codomains of the components of the state function (in this case only two).

DEFINITION 3.11 Let $X_m = \langle M, \mathbb{F}\rangle$ be a functional model for a thing X, where $\mathbb{F}: \langle F_1, F_2, \ldots, F_n\rangle: M \to V_1 \times V_2 \times \ldots \times V_n$ is the state function, and call $\mathbb{L}(X)$ the set of all law statements of X. Then the subset of the codomain $V = V_1 \times V_2 \times \ldots \times V_n$ of $\mathbb{F}$ restricted by the conditions (law statements) in $\mathbb{L}(X)$ is called the *lawful state space* of X in the representation X_m, or $S_{\mathbb{L}}(X)$ for short:

$$S_{\mathbb{L}}(X) = \{\langle x_1, x_2, \ldots, x_n\rangle \in V_1 \times V_2 \times \ldots \times V_n \mid \mathbb{F} \text{ satisfies jointly every member of } \mathbb{L}(X)\},$$

and every point of $S_{\mathbb{L}}(X)$ is called a *lawful* (or *really possible*) *state* of X in the representation X_m.

Clearly, the lawful state space is included in the corresponding conceivable state space: $S_{\mathbb{L}}(X) \subseteq S(X)$ for every $X \in \Theta$ and in any fixed representation X_m of X. For example, if $\mathbb{F} = \langle F_1, F_2\rangle$ takes values in $\mathbb{R}^2$, subject to the restriction (law statement) $\ulcorner F_2 = kF_1 \urcorner$, where k is a real number, then $S_{\mathbb{L}}$ is a one dimensional subspace of $S = \mathbb{R}^2$, i.e. a curve on the plane.

Example In Example 5 (electrostatics), of Sec. 2.4 the conceivable state space was

$$S=\{\langle\rho(x), \varphi(x)\rangle\in\mathbb{R}^2|x\in V\subseteq E^3\}.$$

Since the two components of the state function $\mathbb{F}=\langle\rho, \varphi\rangle$ are tied together by the law statement: $\ulcorner\nabla^2\varphi(x)=4\pi\rho(x)\urcorner$, the lawful state space of the thing is the subspace of S defined by

$$S_{\mathbb{L}}=\{\langle\rho(x), \varphi(x)\rangle\in\mathbb{R}^2|x\in V\subseteq\mathbb{R}^3\ \&\ \nabla^2\varphi(x)=4\pi\rho(x)\}.$$

Because most of the state functions occurring in scientific theories are real valued, and because $\mathbb{R}$ and any of its cartesian powers are vector spaces, the majority of the conceivable state spaces are vector spaces. However, because of the restrictions imposed by the law statements, the lawful state spaces, though usually included in vector spaces, need not be vector spaces themselves – let alone metric spaces.

We close this subsection with a handful of assorted remarks.

Remark 1 If we pretend that the collection of states of a thing is finite, then we are dealing with a *finite state model* of the thing, otherwise with an *infinite state model* of it. Actually there are no finite state things but only model things restricted to a finite number of states, such as automata. The finite state models are reasonable only when our interest is restricted to global properties. In such cases one can use the notion of *capacity* of a thing, defined as $I=\log_2 n$, where n is the number of states of the thing. *Remark 2* The aggregate (or noninteracting composition) of two finite state things can be represented by a state space with $n_1\cdot n_2$ points, where n_1 and n_2 are the number of states of components 1 and 2. Accordingly the capacity of such an aggregate equals the sum of the partial capacities:

$$I(1\dotplus 2)=\log_2(n_1\cdot n_2)=\log_2 n_1+\log_2 n_2=I_1+I_2.$$

But this case, central to linear system theory (cf. Zadeh and Desoer 1963), is too particular to constitute a general basis for ontology. Think that the simplest atom can be in any of infinitely many energy states or in a linear combination thereof. *Remark 3* The next case in order of complexity is that of a thing with an infinite but denumerable state space, such as the collection of stationary states of an atom. In the simplest case of this kind a thing is characterized by a quadruple $\langle n, l, m, s\rangle$ of quantum numbers (radial, orbital, azimuthal, and spin). Beyond a certain energy threshold, though, the integer n must be replaced by a continuous parameter, so that the state space acquires a continuous component. *Remark 4* If, during a part of the existence of a thing (e.g. when it is being observed), only some components of its state function change their values, one says that the remaining components are *ignorable* and the study of the thing can be restricted to the state subspace spanned by the active state variables. This subspace may be called the *reduced state space*. Experimental science presupposes that such a reduction can always be effected. *Remark 5* While in most of general systems theory the state space is assumed to be finite dimensional, the state (or Hilbert) spaces occurring in the quantum theories are infinite dimensional. Every point in such spaces can be analyzed into infinitely many components, namely those along the axes constituted by the orthonormal eigenfunctions of an arbitrary hermitian operator in the Hilbert space.

2.6. *Equivalent Representations of States*

Recall that every law statement for a given thing may be regarded as the value of a certain function L which we called the corresponding *law function* (Sec. 2.3). For a fixed thing the arguments of this function are components of the state function or functions thereof. Therefore a law function may also be construed as a function transforming the state space into itself. That is, for any given thing, $L: S \rightarrow S$ is a function that assigns each state $s \in S_{\mathbb{L}} \subseteq S$ another state $L(s) \in S_{\mathbb{L}}$, not of course an arbitrary one but a state lawfully related to the former. Consequently the totality of laws of a given thing is contained in the set of all the transformations of its state space. This point will be amply exploited in Ch. 5 on change.

Surely not every transformation of a state space represents a law. For one thing the identity map on $S_{\mathbb{L}}$ does not represent any law. Besides, certain transformations of $S_{\mathbb{L}}$, though lawful or law abiding, represent not laws but just different choices of state functions or representations. Consider a certain state function transformation concerning a fixed thing: $\mathbb{F}^* = f(\mathbb{F})$, where f is a function subject to certain restrictions. (A typical case is that f be one-to-one bicontinuous, i.e. that its Jacobian be non-null.) Every such transformation ensues in a different representation of the states of the thing, i.e. in a different state space $S_{\mathbb{L}}^*$. In other words, if $f: S_{\mathbb{L}} \rightarrow S_{\mathbb{L}}^*$ is such a bijection then $f(S_{\mathbb{L}})$ is an alternative state space for the same thing: see Figure 3.2. Again: $X_m = \langle M, \mathbb{F} \rangle$ and $X'_m =$

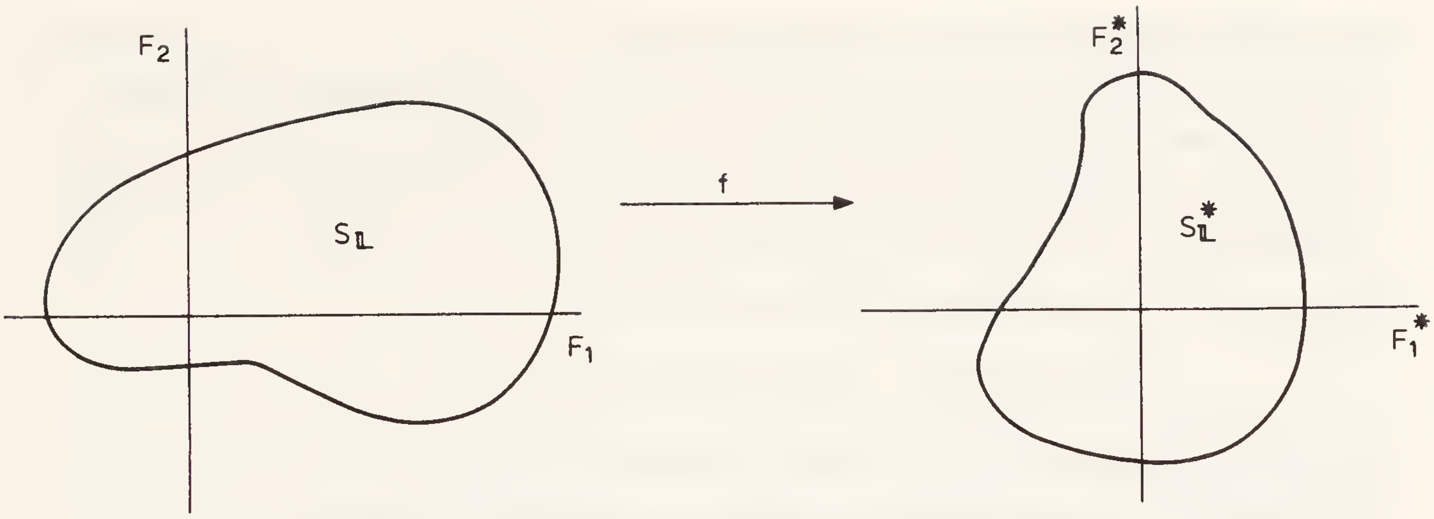


Fig. 3.2. Alternative state spaces for one and the same thing. Two such alternatives are equivalent just in case the law statements in $\mathbb{L}$ are invariant under the bijection f relating the two state spaces.

$=\langle M, f(\mathbb{F})\rangle$ are *alternative representations* of the same thing X iff f is a bijection on $S_{\mathbb{L}}$, where $S_{\mathbb{L}}$ is the space spanned by $\mathbb{F}$ with the restrictions $\mathbb{L}$.

Not every representation change is admissible: only those changes are admissible that keep the laws unchanged. We consecrate this important dichotomy by introducing

DEFINITION 3.12 Let $\mathbb{F}$: $M \to V$ be a state function for a given thing X in a given functional schema $X_m = \langle M, \mathbb{F}\rangle$. Then

(i) the *lawful transformations* of $\mathbb{F}$ are those transformations f: $V \to V^*$ such that $\mathbb{F}^* = f \circ \mathbb{F}$ and that leave the laws $\mathbb{L}$ of X form-invariant;

(ii) every transformed lawful state space $S_{\mathbb{L}}^* = f(S_{\mathbb{L}})$ is an *equivalent representation* of the states of X iff it is the image of $S_{\mathbb{L}}$ under a lawful transformation f;

(iii) the collection of all the equivalent representations of the states of X constitutes *the representation* of the states of X.

Example 1 In special relativistic theories the position and time coordinates transform according to the Lorentz formulas. These are not only laws but induce transformations of any state function such that the basic laws – e.g. the axioms of mechanics or of electrodynamics – retain their form, or are *Lorentz covariant.* And the components of the state function that do not change under a Lorentz transformation are called *invariant.* Examples of invariants: electric charge and entropy. *Example 2* In hamiltonian theories – whether they occur in physics or in biology or elsewhere – the state variables are the generalized coordinates q and momenta p (one set of each for every component of the system of interest). These variables span the state (or phase) space of the thing. New state functions $\mathbb{F}^* = \langle f(q, p), g(q, p)\rangle$ can always be introduced. But only those will be admissible that leave the canonical (or Hamilton's) equations invariant, i.e. those for which $q^* = f(q, p)$ and $p^* = g(q, p)$ are such that

$$\dot{q}^* = \partial H^*/\partial p^* \quad \text{and} \quad \dot{p}^* = -\partial H^*/\partial q^*.$$

Such transformations are called *canonical.* Every one of them induces an equivalent representation of the states of the thing concerned.

2.7. *State and State Preparation*

Whether a thing is in a given state (relative to some frame) as a result of its inner dynamics and its interactions with other things, or as a result of a sequence of human operations, is immaterial for the elucidation of the concept of an objective (though relative) state. After all, the manipulations leading to the preparation of a thing in a given state are strictly physical (or chemical or biological or social) processes even when they are controlled by a supposedly intelligent organism, and they are no less so when the thing concerned happens to be another organism – who is, say, being trained or brainwashed to behave in a certain way.

The previous platitude had to be stated because certain contemporary varieties of subjectivism claim that, unlike macrothings, microthings cannot be regarded as leading an autonomous existence because they are always prepared by some observer to be in (or jump into) preassigned states. More precisely, some quantum mechanics experts have argued as follows. (i) A microthing *is* in no definite state as long as it has not been forced into some state by some observer. (ii) Now, a state preparation is a human operation. (iii) Hence a microthing adopts a definite state only as a result of its being subject to certain human actions. The key formulas that illustrate this argument are the following. Before measuring the property represented by the operator Q, the system is in a superposition of states

$$\psi = \sum c_n\varphi_n, \quad \text{with} \quad Q\varphi_n = q_n\varphi_n,$$

where the φ_n are the eigenfunctions and the q_n the corresponding eigenvalues of the operator Q. Upon measuring whatever Q represents, ψ collapses into one of its components, say φ_m, with probability $|c_m|^2$.

The thing is only now in a definite Q-state φ_m, whereas before the measurement operation (which involves the design and reading of measuring instruments) it was in no definite (eigen)state.

The previous argument is a version of the old subjectivist thesis that properties, hence states, are dependent upon some observer rather than being objective. Moreover, it is not an argument proper, as the conclusion is just a reformulation of the premises (cf. Bunge, 1973b). Besides, the first premise happens to be false: from the fact that I happen not to know the state a thing is in, it does not follow that the thing is in no state at all – unless of course Berkeley's principle *esse est percipii* is adopted. The subjectivist thesis is moreover mathematically untenable. In fact properties are represented by functions (or by operators) and these concepts are not defined unless their entire ranges (of possible values) are given. Whether or not such a range is being sampled by empirical means is irrelevant to the objective (though relative) state of affairs, and so is the circumstance that a microthing may not be in a state that happens to be an eigenstate of the operator we may pick: a state is a state even when it is not an eigenstate of some operator. In other words, for the sake of mathematics one has to assume that the function has all of the values in its range even though only a few of them may be accessible to measurement. Consequently every thing is supposed to be in some state or other even if we have not prepared it to be in any one particular state. When an experimenter prepares a thing he picks a certain subset (occasionally even a point) of the state space of the thing. He thereby induces a real change in a real state of a real thing: he does not create either out of nothing. In sum, despite the subjectivist interpretation of quantum mechanics, states do not depend wholly on the way one looks at them but are objective (though relative) properties of real things.

2.8. *Concluding Remarks*

Every state is a state of some concrete object or other: there are no states in themselves. And conceptual objects are in no states whatsoever. Therefore a thing could be defined as *whatever is in some state or other.* Things differ by the states they are in, and their changes are changes of state. But all the things of a given (natural) kind share the same (lawful) state space – which is a way of saying that they share the same general properties. In sum, states serve to characterize not only individual things but also natural classes of things – hence the centrality of the state concept. But the matter of kinds deserves a special section.

3. FROM CLASS TO NATURAL KIND

3.1. *Classes of Things*

In this section we shall elucidate the concepts of class, kind, and species (or natural kind) of things. To this end we shall make use of some of the ideas introduced in the theory of properties expounded in Ch. 2, Sec. 3 and shall borrow freely from Bunge and Sangalli (1977). To begin with the scope function $\mathcal{S}$ introduced by Definition 2.5 of Ch. 2, Sec. 3.2 can easily be redefined for things:

DEFINITION 3.13 The *scope* of a property is the set of things possessing it. I.e. the function $\mathcal{S}: \mathbb{P} \rightarrow 2^{\Theta}$ such that '$x \in \mathcal{S}(P)$', for $x \in \Theta$ and $P \in \mathbb{P}$, is interpretable as "*x possesses P*", is the *scope.*

Correspondingly the Definition 2.6 of a class of substantial individuals now reads

DEFINITION 3.14 A subset X of the set Θ of things is called a *class* (of things) iff there is a property $P \in \mathbb{P}$ such that $X = \mathcal{S}(P) \in 2^{\Theta}$.

And Postulate 2.6 now reads

POSTULATE 3.5 The intersection of any two classes of things, if nonempty, is a class: For any two compatible properties $P, Q \in \mathbb{P}$ there is at least a third property $R \in \mathbb{P}$ such that $\mathcal{S}(R) = \mathcal{S}(P) \cap \mathcal{S}(Q)$.

We proceed to investigate the algebra of classes of things, which differs from the algebra of sets if only because unions and complements are not defined in our theory. A useful tool for this investigation is the modified scope function introduced by Definition 2.12(iii) of Ch. 2, Sec. 3.4. In terms of the thing concept, it is the function $[\mathcal{S}]: [\mathbb{P}] \rightarrow 2^{\Theta}$ that assigns each set of concomitant properties the collection of things possessing them.

3.2. *Ideals and Filters of Classes of Things*

The modified scope function $[\mathcal{S}]: [\mathbb{P}] \rightarrow 2^{\Theta}$, which is injective, induces whatever structure its codomain has on its domain. Now, the power set of any nonempty set, such as 2^{Θ}, has the lattice structure under the inclusion relation. Indeed, 2^{Θ} is closed under the union and intersection operations, and intersections precede (are included in) unions. So, $\langle 2^{\Theta}, \cup, \cap, \subseteq \rangle$ is a lattice. And, being a lattice, we can define ideals and filters on it. Since these structures will turn out to have interesting ontological interpretations, we shall briefly recall the notions of an ideal and a filter.

An *ideal* of a lattice $\mathscr{L} = \langle L, \vee, \wedge, \leqslant \rangle$ is a structure $\mathscr{I} = \langle I, \vee, \leqslant \rangle$ such that

(i) $I \subseteq L$;
(ii) every ancestor is in: if $x \in L$ and $y \in I$ and $x \leqslant y$, then $x \in I$;
(iii) all suprema are in: if $x, y \in I$, then $x \vee y \in I$.

We designate by $\mathscr{I}\mathscr{L}$ the collection of all the ideals of $\mathscr{L}$. This set has a lattice structure. If $I, I' \in \mathscr{I}\mathscr{L}$, their infimum is $I \cap I'$, while their supremum, designated $I \sqcup I'$, is the intersection of all the ideals which contain both I and I'.

Dually, a *filter* of a lattice $\mathscr{L} = \langle L, \vee, \wedge, \leqslant \rangle$ is a structure $\mathscr{F} = \langle F, \wedge, \leqslant \rangle$ such that

(i) $F \subseteq L$;
(ii) every successor is in: if $x \in F$ and $y \in L$ and $x \leqslant y$, then $y \in F$;
(iii) all infima are in: if $x, y \in F$, then $x \wedge y \in F$.

We call $\mathscr{F}\mathscr{L}$ the collection of all filters of $\mathscr{L}$. $\mathscr{F}\mathscr{L}$ is a lattice.

It will soon become apparent that neither entities nor their properties are stray: substantial individuals come in filters while substantial properties come in ideals.

We introduce next a sort of inverse of the scope function, namely a function that assigns each family of sets of entities a collection of properties. More precisely, look at

DEFINITION 3.15 The *starred scope* is the function from the power set of the power set of the collection of things into the power set of the collection of properties,

$$\mathscr{S}^*: 2^{2^{\Theta}} \to 2^{\mathbb{P}},$$

such that $\mathscr{S}^*(\mathscr{X}) = \{P \in \mathbb{P} \mid \mathscr{S}(P) \in \mathscr{X}\}$, where $\mathscr{X} \subseteq 2^{\Theta}$ is a family of subsets of Θ.

It will be recalled that, since $\mathscr{S}(P \wedge Q) = \mathscr{S}(P) \cap \mathscr{S}(Q)$, the scope function takes suprema into infima. Well, the starred scope function does just the opposite: if $\mathscr{X}$ is a filter of subsets of Θ then $\mathscr{S}^*(\mathscr{X})$ is an ideal of properties. Hence $\mathscr{S}^*$ can be construed as a function

$$\mathscr{S}^*: \mathscr{F}(2^\Theta) \to \mathscr{I}\mathbb{P}$$

that takes filters of sets of entities into ideals of substantial properties.

Our next step requires

DEFINITION 3.16 If $T \subseteq \Theta$ is a set of things, then the *filter of* 2^Θ *generated by* T is the intersection of all the filters of 2^Θ that count T as a member, i.e. $[T) = \bigcap\{X \in \mathscr{F}(2^\Theta) \mid T \in X\}$, or also $[T) = \{X \in 2^\Theta \mid T \subseteq X\}$.

If we now apply the definition of $\mathscr{S}^*$ to the computation of $\mathscr{S}^*([T))$, we find

$$\begin{aligned}\mathscr{S}^*([T)) &= \{P \in \mathbb{P} \mid \mathscr{S}(P) \in [T)\} = \{P \in \mathbb{P} \mid T \subseteq \mathscr{S}(P)\} \\ &= \{P \in \mathbb{P} \mid \text{For all } x \in \Theta\text{: } x \in T \Rightarrow x \text{ possesses } P\}.\end{aligned}$$

Thus $\mathscr{S}^*([T))$ turns out to be the set of all properties possessed by all members of T. But, by the remarks following Definition 3.15, $\mathscr{S}^*$ takes its values in the ideals of P. Hence $\mathscr{S}^*([T))$ is an ideal of P. Hence by the Definition 2.3 of p, we obtain

COROLLARY 3.3 Let $x \in T \subseteq \Theta$. Then

(i) $p(T) = \mathscr{S}^*([T)) =$ The ideal of all the properties of all T's;

(ii) $p(x) = \mathscr{S}^*([x)) =$ The ideal of all the properties of thing x.

To put it another way: Both the set of properties of a thing and the set of all the properties shared by all the members of an arbitrary set of things, far from being unstructured sets, are ideals. In particular if the set of things happens to be a class then the set of all the properties shared by all the members of the class consists of all the properties preceding the property defining the class. In fact we have

THEOREM 3.1 Let $T = \mathscr{S}(P)$ be the class defined by property $P \in \mathbb{P}$. Then the collection of properties of the members of T equals the set of all the properties preceding P:

$$p(T) = \{Q \in \mathbb{P} \mid Q \preccurlyeq P\}.$$

Proof For all $Q \in \mathbb{P}$, $T \subseteq \mathscr{S}(Q)$ iff $\mathscr{S}(P) \subseteq \mathscr{S}(Q)$ iff $Q \preccurlyeq P$.

Note that $p(T)$ is the *principal ideal* of P in $\mathbb{P}$, i.e. $(P)_{\mathbb{P}}$. (See Figure 3.3.) Note also that $p(T)$ is not an arbitrary notation but the value of the

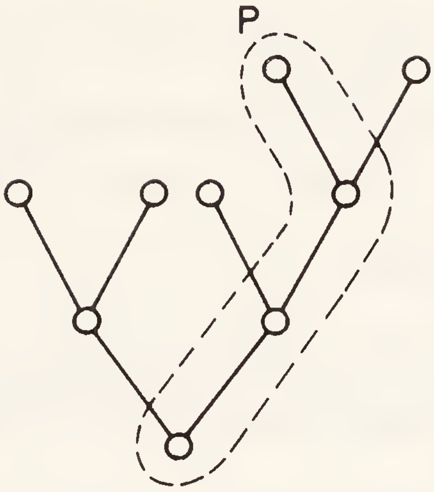


Fig. 3.3. The principal ideal of a property P is a fragment of the tree of properties. The dotted lines indicate $(P)_{\mathbb{P}}$, the bottom element represents a universal property.

function $p: 2^{\Theta} \rightarrow \mathscr{I}\mathbb{P}$ at $T \in 2^{\Theta}$. This function p reverses order, in the sense that the greater the set of things the fewer properties they share. In fact we have

COROLLARY 3.4 Let $T, T' \subseteq \Theta$. Then

$$\text{If } T \subseteq T' \text{ then } p(T') \subseteq p(T).$$

3.3. *Kinds and Species*

Whereas a single property determines a class (Definition 3.14), a set of properties will be said to determine a *kind*, and a set of lawfully related properties a *natural kind*. That is, the members of a kind are all those and only those things that share all the properties in the given set. (See Figure 3.4.)

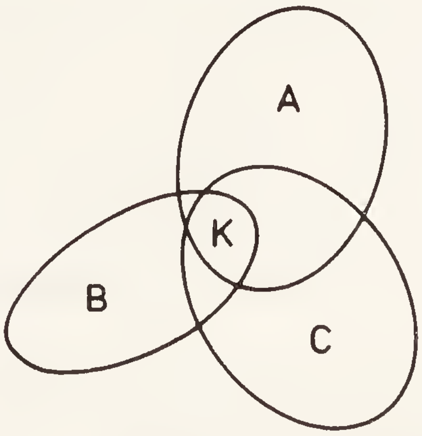


Fig. 3.4. Kind K is the intersection of classes A, B, and C.

More precisely, we make

DEFINITION 3.17 Let $k: 2^{\mathbb{P}} \to 2^{\Theta}$, be the function assigning to each nonempty set $\mathbb{R} \in 2^{\mathbb{P}}$ of substantial properties the set

$$k(\mathbb{R}) = \bigcap_{P \in \mathbb{R}} \mathscr{S}(P)$$

of things sharing the properties in $\mathbb{R}$. This value $k(\mathbb{R})$ is called the $\mathbb{R}$-*kind* of things. [Throughout this subsection and the next, $\mathbb{R}$ does not designate the real line.]

If $\mathbb{R}$ is finite, the corresponding $\mathbb{R}$-kind is a class. In fact calling $\mathbb{R} = \{P_1, P_2, \ldots, P_n\}$,

$$k(\mathbb{R}) = \bigcap_{i=1}^{n} \mathscr{S}(P_i) = \mathscr{S}(P_1 \wedge P_2 \wedge \ldots \wedge P_n),$$

so that $k(\mathbb{R})$ is the class of entities possessing the single (though compound) property $P_1 \wedge P_2 \wedge \ldots \wedge P_n$.

The notion of an $\mathbb{R}$-kind underlies all classing. In fact it allows one to introduce the notion of equivalence, or equality in certain respects, which is the very basis for the grouping of items:

DEFINITION 3.18 Let $\mathbb{R}$ be a set of substantial properties. Then two things $x, y \in \Theta$ are said to be $\mathbb{R}$-*equivalent*, or *equal in every respect* $P \in \mathbb{R}$, iff they possess precisely the same properties in $\mathbb{R}$:

$$x \sim_{\mathbb{R}} y =_{df} (P)(P \in \mathbb{R} \Rightarrow (x \text{ possesses } P \Leftrightarrow y \text{ possesses } P))$$

or, equivalently,

$$x \sim_{\mathbb{R}} y =_{df} p(x) \cap \mathbb{R} = p(y) \cap \mathbb{R}.$$

For example, since all the cells in an organism originate in a single cell, they are genetically identical. Even being different in other respects they are equal in the genetic respect, or G-equivalent. The totality of such cells, *qua* set not *qua* entity (organism), is called a *clone.*

Since $\sim_{\mathbb{R}}$ is an equivalence relation on S, it splits any set T of things into disjoint equivalence classes, variously called species, genera, orders, and so on. Such a splitting (or grouping) deserves a name:

DEFINITION 3.19 Let $\mathbb{R} \subset \mathbb{P}$ be a set of substantial properties and $T \subset \Theta$ a set of things. Then the $\mathbb{R}$-*classing* of T is the partition $T/\sim_{\mathbb{R}}$, and every member of it is called a *cell.*

Different equivalence relations induce different partitions. Of particular interest are those pairs of classings that, far from being on the same footing, are of different degrees of fineness (or coarseness). This notion is made precise by

DEFINITION 3.20 Let $\mathbb{R}$ and $\mathbb{R}'$ be sets of substantial properties, $\sim_{\mathbb{R}}$ and $\sim_{\mathbb{R}'}$ their respective equivalence relations, and $T/\sim_{\mathbb{R}}$ and $T/\sim_{\mathbb{R}'}$ their respective classings. Then $\sim_{\mathbb{R}}$ and $T/\sim_{\mathbb{R}}$ are *finer* than $\sim_{\mathbb{R}'}$ and $T/\sim_{\mathbb{R}'}$ respectively iff $\mathbb{R}' \subseteq \mathbb{R}$, in which case, for any $x, y \in S$,

(i) If $x \sim_{\mathbb{R}} y$ then $x \sim_{\mathbb{R}'} y$;

(ii) each cell of $T/\sim_{\mathbb{R}'}$ is the union of cells of $T/\sim_{\mathbb{R}}$.

This is precisely how the classings of items into species, genera, etc., are performed, namely with the help of equivalence relations of different power. (The higher powered of two equivalence relations is the finer. And the relation of fineness, whereby fine precedes coarse, is a partial order.)

The number of cells contained in each partition $T/\sim_{\mathbb{R}}$ depends upon $\mathbb{R}$, i.e. upon the number of respects that are being considered. Since there are at most as many $\sim_{\mathbb{R}}$-cells as there are subsets of $\mathbb{R}$, the following important result is obtained:

COROLLARY 3.5 If $\mathbb{R}$ is finite, so is the corresponding partition $T/\sim_{\mathbb{R}}$ of an arbitrary set T of entities.

Since in practice only a finite number of properties is handled, every classing of any collection of entities results in a finite number of classes. One of these classes is the scope of the conjunction of all the properties in $\mathbb{R}$.

Now, there are two ways of ascertaining whether two things are in the same $\sim_{\mathbb{R}}$-cell or are $\sim_{\mathbb{R}}$-equivalent. One is to note and match all their observable properties regardless of their weight: this is the pretheoretical method characteristic of classical taxonomy and also the one adopted by contemporary numerical taxonomy. It can be misleading, for a uniform appearance may hide important differences – and small differences may conceal basic kinship.

An alternative method, and the deepest, consists in grouping things by the laws they possess. When laws are made the *fundamentum divisionis* of a set of things, the resulting kinds are maximally natural – or, in Aristotelian jargon, accident is then unlikely to prevail over essence. The outcome is a set of natural kinds or species. Such sets – such as the chemical species and the biological species characterized in evolutionary terms – are typical of advanced modern science in contrast to purely empirical knowledge and to descriptive science. The definition of the concept of natural kind is obtained by just specializing that of an $\mathbb{R}$-kind (Definition 3.17). More precisely, we can get it by taking the restriction of the kind function k to the power set of the collection of all laws:

DEFINITION 3.21 Let $\mathbb{L}$ be the set of laws and let $k_{\mathbb{L}}: k|2^{\mathbb{L}}: 2^{\mathbb{L}} \to 2^{\Theta}$ be the function assigning to each $\mathbb{L}_i \subset \mathbb{L}$ of laws the set

$$k(\mathbb{L}_i) = \bigcap_{L \in \mathbb{L}_i} \mathcal{S}(L)$$

of entities sharing the laws in $\mathbb{L}_i$. This value $k(\mathbb{L}_i)$ is called the $\mathbb{L}_i$-*species* or *natural kind.*

Things have then properties of two kinds: the laws characteristic of their natural kind, and idiosyncratic properties: see Figure 3.5. That is, we can make

DEFINITION 3.22 Let $\mathbb{L}_i \subset \mathbb{L}$ be a set of laws and $k(\mathbb{L}_i)$ the corresponding natural kind. Further, call $p(x)$ the set of properties of a member x of $k(\mathbb{L}_i)$. Then

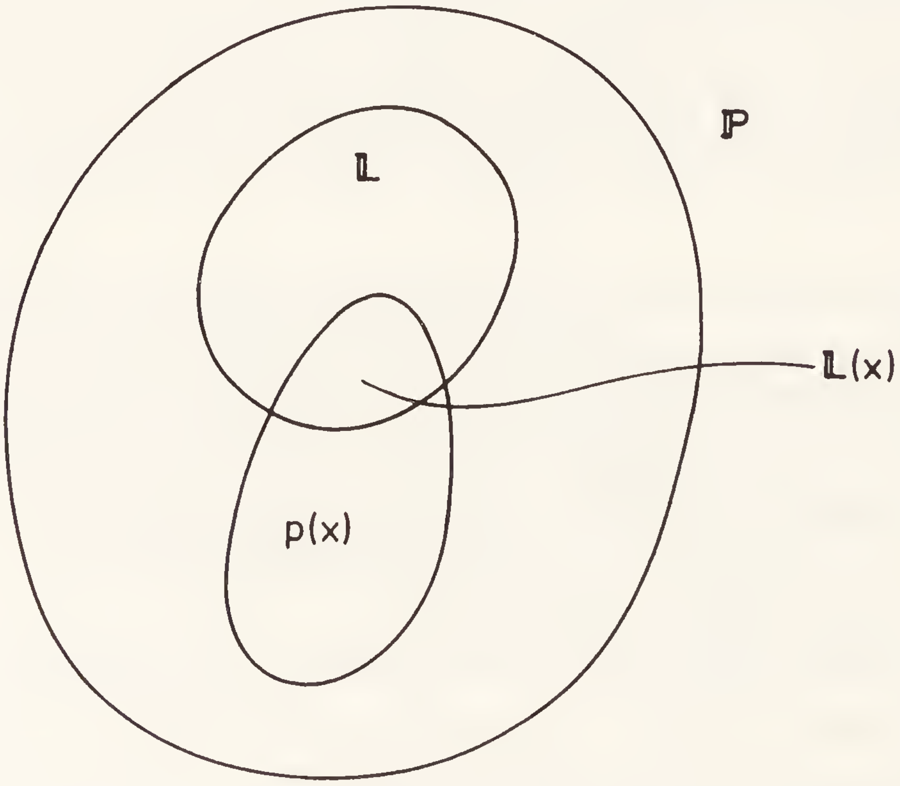


Fig. 3.5. Some of the properties of a thing x are its laws $\mathbb{L}(x)$; the rest, $p(x) - \mathbb{L}(x)$, are its idiosyncratic properties.

(i) the *species-specific properties* of x are all those in $\mathbb{L}_i$;

(ii) the *idiosyncratic properties* of x constitute the complement of its species-specific properties, i.e. $p(x)-\mathbb{L}_i$.

Two things sharing the properties characteristic of their common kind, though possibly different in every other respect, are equivalent with respect to the laws defining their kind. For example two chunks of material are golden iff, whatever their shape, weight, location, etc., they satisfy the laws defining the natural kind (chemical species) Au. In general:

DEFINITION 3.23 Let x and y be two things and $\mathbb{L}_i \subset \mathbb{L}$ a set of laws. Then x and y are *nomologically equivalent relative to* $\mathbb{L}_i$ iff x and y share all their species-specific properties, i.e. all the laws in $\mathbb{L}_i$:

$$x \sim_{\mathbb{L}_i} y =_{df} p(x) \cap \mathbb{L}_i = p(y) \cap \mathbb{L}_i.$$

And of course two things are *nomologically inequivalent* if they do not possess exactly the same laws – or, what amounts to the same, if they do not belong to the same natural kind. Hence just as individuation consists in the emergence (or the recognition) of a thing with idiosyncratic properties, so speciation consists in the emergence (or the recognition) of a collection of things with peculiar laws – i.e. a natural kind or species.

Finally we combine the notion of natural kind with that of composition (Definition 1.6). An entity may or may not be composed of parts of the same natural kind. If it is not so composed we say it is an atom of its kind. More precisely we propose

DEFINITION 3.24 Any thing belonging to a natural kind $k(\mathbb{L}_i)$ is an *atom* of that kind, or $k(\mathbb{L}_i)$*-elementary*, iff it has no proper parts of the same kind. Otherwise it is called *molecular*.

3.4. *The Algebra of Kinds*

Since properties do not satisfy the algebra of predicates – i.e. they do not form Boolean algebras – it may be surmised that kinds do not obey the algebra of sets. The following investigation will show that this is indeed the case.

Recall how the kind function k was introduced by Definition 3.17. Like p, the kind function k reverses order:

$$\text{If } \mathbb{R} \subseteq \mathbb{R}' \text{ then } k(\mathbb{R}') \subseteq k(\mathbb{R}).$$

I.e., as more conditions are added fewer individuals satisfy them. We have then two order reversing functions relating the family of sets of properties and the family of sets of things:

$$2^{\mathbb{P}} \underset{p}{\overset{k}{\rightleftarrows}} 2^{\Theta}.$$

While k assigns to every $\mathbb{R} \subseteq \mathbb{P}$ the set of all things possessing every one of the properties in $\mathbb{R}$, p assigns to a subset T of Θ the set of all the properties shared by all members of T. More precisely we find the following bridge between sets of properties and sets of things:

THEOREM 3.2 For all $\mathbb{R} \subseteq \mathbb{P}$ and all $T \subseteq \Theta$,

$$\mathbb{R} \subseteq p(T) \text{ iff } k(\mathbb{R}) \supseteq T.$$

Proof If $\mathbb{R} \subseteq p(T)$ then, for all $Q \in \mathbb{R}$, $T \subseteq \mathscr{S}(Q)$. So, $T \subseteq \bigcap\{\mathscr{S}(Q) \mid Q \in \mathbb{R}\} = k(\mathbb{R})$. Conversely, if $T \subseteq k(\mathbb{R}) = \bigcap\{\mathscr{S}(Q) \mid Q \in \mathbb{R}\}$, then $T \subseteq \mathscr{S}(Q)$ for all $Q \in \mathbb{R}$, so $\mathbb{R} \subseteq p(T)$.

This theorem has a number of interesting consequences – first of all

THEOREM 3.3 The composition pk (i.e. function k followed by function p) is a closure operator on $2^{\mathbb{P}}$. I.e., for all $\mathbb{R}, \mathbb{R}' \subseteq \mathbb{P}$

(i) $\mathbb{R} \subseteq pk(\mathbb{R})$;
(ii) $\mathbb{R} \subseteq \mathbb{R}' \Rightarrow pk(\mathbb{R}) \subseteq pk(\mathbb{R}')$;
(iii) $pk\,pk(\mathbb{R}) = pk(\mathbb{R})$.

Proof (i) follows from Theorem 3.2 by taking $T = k(\mathbb{R})$. (ii) follows from the fact that both k and p reverse order. Proof of (iii): $pk(\mathbb{R}) \subseteq pk\,pk(\mathbb{R})$ follows from (i). To check the inclusion the other way use first Theorem 3.2 with $\mathbb{R} = p(T)$ to get: For all $T \subseteq \Theta$, $kp(T) \supseteq T$. Now apply twice this last result, with $T = k(\mathbb{R})$ and $T = kp(k(\mathbb{R}))$, obtaining $k(\mathbb{R}) \subseteq kp(k(\mathbb{R})) \subseteq kp\,kp(k(\mathbb{R}))$. Hence by Theorem 3.2 $pk\,pk(\mathbb{R}) \subseteq pk(\mathbb{R})$.

Because of the symmetry described by Theorem 3.2 there is the further consequence that kp too is a closure operator on 2^{Θ}.

The subsets $\mathbb{R}$ of $\mathbb{P}$ such that $pk(\mathbb{R}) = \mathbb{R}$ will be said to be *closed* with respect to pk. By Theorem 3.1 only subsets of $\mathbb{P}$ which are ideals can be closed. We shall see in the sequel that these closed ideals play a special role. They are precisely those sets $\mathbb{I}$ of properties such that the things that share all the properties in $\mathbb{I}$ do not share any other properties. I.e. in general we have clause (i) of Theorem 3.3, which just states that, for a given $\mathbb{I}$, the things sharing all the properties in $\mathbb{I}$ (viz., $k(\mathbb{I})$) may also share other properties: $\mathbb{I} \subseteq pk(I)$. But if $\mathbb{I} = pk(\mathbb{I})$, then $\mathbb{I}$ is a set of properties characterized by the set of individuals which possess them.

We arrive thus at the next consequence of Theorem 3.2, which supplies alternative characterizations of the concept of a kind:

THEOREM 3.4 For any subset $T \subseteq \Theta$ of things the following statements are equivalent:

(a) T is a kind (i.e. there exists a subset $\mathbb{R}$ of $\mathbb{P}$ such that $T = k(\mathbb{R})$).

(b) T is determined by the properties shared by all of its members (i.e. $T = k(p(T))$).

(c) T is an $\mathbb{I}$-kind for a (unique) closed ideal $\mathbb{I}$ (i.e. there exists a necessarily unique closed ideal $\mathbb{I}$ of $\mathbb{P}$ such that $T = k(\mathbb{I})$).

Proof (a)$\Rightarrow$(b): apply kp to both members of $T = k(\mathbb{R})$ to get $kp(T) = kpk(\mathbb{R})$. Using Theorem 3.2 check that $kpk(\mathbb{R}) = k(\mathbb{R})$.

(b)$\Rightarrow$(c): it suffices to show that $p(T)$ is a closed ideal (the uniqueness assertion is easy to check). From the fact that kp is a closure operator we get $T \subseteq k(pkp(T))$. By Theorem 3.2 it now follows that $pkp(T) \subseteq p(T)$.

(c)$\Rightarrow$(a): trivial.

From the last clause of Theorem 3.4 it follows that, in order to generate all kinds, it is enough to consider the closed ideals of properties. Indeed, the correspondence between kinds and closed ideals is one to one. Furthermore we have

THEOREM 3.5 The collection of all kinds forms a complete lattice under inclusion, and this lattice is isomorphic to the dual of the lattice of all closed ideals.

Proof If χ is a family of kinds, the intersection $\bigcap\limits_{X \in \chi} X$ is again a kind and it is the infimum of χ. The supremum is the kind $k\left(p\left(\bigcup\limits_{X \in \chi} X\right)\right)$. The order reversing isomorphism is the function k that sends the closed ideal $\mathbb{I}$ into the $\mathbb{I}$-kind $k(\mathbb{I})$. Indeed, for any two closed ideals $\mathbb{I}$, $\mathbb{I}'$, applying Theorem 3.2 with $\mathbb{R} = \mathbb{I}$ and $T = k(\mathbb{I}')$ we get $\mathbb{I} \subseteq \mathbb{I}'$ iff $k(\mathbb{I}) \supseteq k(\mathbb{I}')$. (The inverse isomorphism is the function p.)

Note the difference between the structures of the family of kinds and the family of subsets of Θ (i.e. the classes). Whereas the latter forms a Boolean algebra, the family of kinds does not: for one thing the union of two kinds is not necessarily a third kind. This is of course a feature of the world and one that a purely extensional or nominalistic approach cannot possibly reveal.

Finally, Theorem 3.5 may be given the following twist. Designate with $\mathbb{P}^*$ the collection of all closed ideals of properties. This set contains not only the substantial (or really possessed) properties but others as well. In this larger set $\mathbb{P}^*$ the sup of arbitrary (possibly infinite) families of elements exist; they can be interpreted as the conjunction of (possibly infinite) families of properties. Theorem 3.5 tells us then that there is a 1 : 1 correspondence between the kinds of things and the sum total of (real and imaginary) properties. This correspondence is the function k, which may be regarded as a generalization of the scope function $\mathscr{S}$, for it agrees with $\mathscr{S}$ on the set $\mathbb{P}$ of substantial properties. (I.e. $k((P]) = \mathscr{S}(P)$, where $(P]$ is the closed ideal $\{Q \in \mathbb{P} | Q \preccurlyeq P\}$.) Besides, k also assigns scopes $k(\mathbb{I})$ to the property ideals. Moreover the correspondence k is such that to the (possibly infinite) conjunction (i.e. the supremum) of a set of properties corresponds the intersection of their scopes.

So much for the algebra of kinds – which applies of course to natural or nomological kinds as a particular case.

We now have a theory of properties, distinct from the theory of predicates, and a theory of kinds, different from the algebra of sets. We can therefore use without qualms the concepts of a property and a kind. The differences between predicates and properties, and between sets and kinds, suffice to ruin the ontological interpretations of logic and of set theory. There is no reason to expect that pure mathematics is capable of disclosing, without further ado, the structure of reality.

3.5. *Variety*

We close this section by casting a glance at the concepts of variety. There are at least two of them. One consists in the diversity of kinds of thing "represented" (to speak platonically) in a restricted set of things, such as a random sample of the collection of organisms present in one cubic centimeter of top soil. There are several measures of this kind of variety, or *local variety*: the sheer number of different species, various indices of ecological variety, or variety conducive to ecological equilibrium (see Pielou, 1969), the information theoretic measure (see Ashby, 1956), and so on. These are all specific measures of the variety found, or to be expected, in some restricted set of things. Hence their study belongs in science.

Here we are interested in the concept of *universal variety*, or the diversity in the sum total of things. We have assumed that there are infinitely many things (Postulate 3.1). But of course we have empirical access to smallish samples of that totality. However we can transcend this limitation by dint of imagination – mainly by classing and by conjecturing patterns. Classing enables us to play down idiosyncrasies and it prepares the ground for finding regularities. And hypothesizing general patterns, or laws, allows us to play down particular circumstances as well as to refine and deepen the previous classings.

Now, we grasp laws only through our conceptual reconstruction of them, i.e. law statements. Among such scientific formulas there are some which discard both particular circumstances and idiosyncrasies: they are the *generic* or *basic* laws, in contrast to the specific or derivative ones. This difference is logically, methodologically and ontologically crucial and it boils down to this. Whereas every scientific theory contains at best a finite number of generic or basic laws, it also contains all the logical consequences obtained by conjoining such axioms with other premises of the theory – and these consequences are just as entitled to be called 'laws'. The basic/derived, or generic/specific dichotomy is so fundamental that only the basic (or generic) laws are required to possess covariance properties (Bunge, 1967b).

The consequence for the matter of universal variety is rather obvious, namely this. Whereas classing things according to any laws guarantees infinite variety (because there are infinitely many law statements), classing according to basic laws alone warrants the partition of the set of things into a finite number of natural kinds. These two modes of classing are so important that they warrant one more convention:

DEFINITION 3.25 The set of all things that share a basic law is called a *natural genus*, and that which shares a particular law, a *natural species*.

This law-natural kind correspondence, together with our metatheoretical knowledge of the structure of scientific theories, entails

THEOREM 3.6 There is a finite number of natural genera and there are infinitely many natural species.

We realize now that the question 'How varied is reality?' is ambiguous. The answer is: Whereas the specific variety of reality is infinite, its generic variety is finite. And, whereas the latter trait of reality renders scientific research possible, the former makes it open ended.

4. THE WORLD

4.1. *What Does the World Consist in and of?*

Most metaphysicians have asked, and some have answered, the following questions: (i) What does the world consist in – i.e. what basic kinds of thing are its constituents?, and (ii) What does the world consist of – i.e. what is there? We have an answer to the first question not to the second.

Our answer to the question about the kinds of thing the world is "made" of follows from Postulate 3.2 and Definition 3.5: *The world is the aggregation of its constituents, which are things.* Or, if preferred, the world is that thing which is the physical sum of all concrete or material existents. This is of course a materialist answer. However, most materialists seem to prefer the formula "The world is material". We do not adopt the latter formula because it suggests the physicalist thesis that the only realities are physical systems, while in point of fact living systems and supersystems (e.g. ecosystems and societies) are just as real – and concrete.

Our answer, being neither new nor striking, is not popular among the more sophisticated contemporary philosophers. Indeed consider but a few typical contemporary answers to question (i): The world consists in sensations (Mach, 1906) or in *Erlebnisse* or elementary experiences (Carnap, 1928); the world is the totality of facts (Wittgenstein, 1922); the world is the sum total of processes (Whitehead, 1929); the world is the sum of individuals of all kinds (Goodman, 1956); the world is made of bodies and persons (Strawson, 1959). Having to sketch our own view on the matter we shall not attempt to discuss any of these alternatives. Nevertheless we cannot help but submit that none of those views is continuous with science, hence none of them worthy of belonging in scientific metaphysics. In particular the subjectivist doctrines are inconsistent with natural science; the view that the world consists in facts, or events, or processes, is incompatible with physical cosmology and fails to clarify the very notions of fact, event, and process; the body/person dualism conflicts with physiological psychology and the evolutionary study of behavior; and the thesis that the world consists of individuals is unsatisfactory unless one adds the clause that such individuals are concrete. On the other hand, the thesis that the world is the aggregate of all things, is continuous with science, for things are what every factual science investigates.

As for the second question – namely What is there? – we shall abstain from answering it. That is, we shall not list the kinds of constituent of the world but shall leave the task to the special sciences. For, no sooner does the metaphysician pronounce the world to be "made of" such and such kinds, than the scientist discovers either that some of the alleged species are empty or that others are missing in the metaphysician's list.

4.2. *Individuals, Populations, Communities, and Species*

Scientists do not stop at individuals but group and interrelate them. And they gather things not into arbitrary sets but into kinds and, more particularly, into natural kinds. They recognize as many natural kinds as laws, yet they manage to keep down the count of natural species. This they do, not only by focusing on basic or generic laws (cf. Sec. 3.5) but also, when the latter are not available, by disregarding slight individual differences such as local (ecological) variations which, though real, are rated as secondary – rightly or wrongly so. In other words, scientists distinguish species from varieties and, though keeping an eye on the latter (for some of them may eventually evolve into separate species), they focus on the former. (For the methodological injunction *Species non sunt multiplicanda praeter necessitatem* see Hooker, 1853, Introductory Essay pp. xii–xvii, and Allen, 1870–71.)

Scientists, in particular biologists, recognize three levels of concrete entity: the individual thing, the aggregate of individuals of a given kind, or population, and the aggregate of individuals of different kinds, or mixture of populations (e.g. the ecosystem or biological community). Unlike both an individual thing and a kind, a population is a concrete aggregate of entities of one or more species. In particular a biological population is a group of organisms of a given species inhabiting a limited territory, interacting and interbreeding, and relatively isolated from other populations. According to modern evolution theory, the evolutionary unit – that involved in inheritance and selection – is neither the individual organism nor the species but the population (or perhaps even the ecosystem). This is so because the organisms in a population share a gene pool and an environment. (Different populations of the same species are detached: they share neither a gene pool nor exactly the same environment, and are thus likely to evolve exclusive characters whose accumulation and selection may result in speciation, i.e. the formation of a population belonging to a new species.) A species, on the other hand, has neither a uniform environment nor a common set of genes, hence no selection and survival mechanism: unlike a population, which is a thing, a species is a concept – albeit an indispensable one.

The logical relation between the concepts of individual, population, and species are elucidated by

DEFINITION 3.26 A thing X is

(i) a *monospecific population* iff X is the aggregation of some set A of individual things belonging to a single natural kind $K_{\mathbb{L}}$:

$$X=[A] \quad \text{with} \quad A \subseteq K_{\mathbb{L}};$$

(ii) a *polyspecific population* (or *community* or *mixture* as the case may be) iff X is the aggregation of two or more sets A_i of individuals belonging to different natural kinds $K_{\mathbb{L}_i}$:

$$X=\left[\bigcup_i A_i\right] \quad \text{with} \quad A_i \subseteq K_{\mathbb{L}_i}$$

$$\text{and} \quad K_{\mathbb{L}_i} \neq K_{\mathbb{L}_j} \quad \text{for} \quad i \neq j, 1 \leqslant i, j \leqslant n.$$

(Conversely, any given species or natural kind may be regarded as the union of the membership of all the possible populations of things of the kind. But this construal cannot be formalized in our system since we have defined a population with the help of the species concept.)

The preceding considerations are relevant to overcoming the nominalism/realism deadlock in biological taxonomy. The question is whether biology concerns individual organisms or species of such. Our answer is of course: Both plus populations, for taxonomic systems are relations among natural kinds, and a kind is a collection of individuals which, in the case of organisms, live in populations. Thus we escape being impaled in either horn of the dilemma: we do not have to choose between individuals belonging to no species (nominalism) and species transcending their membership (Platonism). Both individual organisms and populations of such are real, but both share certain clusters of properties (in particular laws) and thus belong to definite natural kinds. The Platonist is right in holding that every individual belongs to ("participates in") some species (or in the union of several species), wrong in assigning priority to the latter: a set consists of its membership. And the nominalist is right in claiming that kinds are things of the mind but wrong in holding that all kinds are arbitrary inventions: natural kinds are inventions adequate to reality insofar as they are sets of nomologically equivalent things.

In sum each of the philosophical poles, nominalism and realism, contains a grain of truth, none the whole truth, and both ignore the *tertium quid* that population genetics and ecology have interpolated between the individual organism and its species, namely the population. This conclusion reinforces our previous critique of Platonism in Ch. 2, Sec. 5.2: there are no universals in and by themselves. There are only things with definite properties. A general property (in particular a law) may be said to be universal in a given kind if every member of the latter possesses that property. (Recall Definition 2.18.)

4.3. *Existence Concepts*

The ordinary language expression 'there are' is ambiguous, as it designates two different concepts: the logical concept *something* and the ontological concept of *existence*. Logic takes care of the former, analyzing it as the existential quantifier, which it might be better to rechristen *particularizer*, or *indefinite quantifier*, to distinguish it from both the *universalizer* (or universal quantifier) and the *individualizer* (or descriptor).

Surely most contemporary philosophers hold that $\exists$ formalizes both the logical concept "some" and the ontological concept of existence. I shall argue that this is a mistake. Consider the statement "Some sirens are beautiful", which can be symbolized "$(\exists x)(Sx \; \& \; Bx)$". So far so good. The trouble starts when the formula is read "There are beautiful sirens". The existential interpretation is misleading because it suggests belief in the real existence of sirens, while all we intended to say was "Some of the sirens existing in Greek mythology are beautiful". The particularizer formalizes the prefix 'some' but not the phrase 'existing in Greek mythology'. (The juxtaposition of two particularizers bearing on the same variable makes no sense.) We need then an exact concept of existence different from $\exists$. Much to the dismay of most logicians we shall introduce one in the sequel. In fact we shall introduce an *existence predicate*, thus vindicating the age-old intuition that existence is the most important property anything can possess.

We shall introduce the concept of relative or contextual existence, exemplified by "Disjunction exists in logic but not in nature", and "Birds exist in nature but not in logic". We do that by making

DEFINITION 3.27 Let A be a well formed set included in some set X, and χ_A the characteristic function of A, i.e. the function $\chi_A : X \rightarrow \{0, 1\}$ such that $\chi_A(x) = 1$ iff x is in A, and $\chi_A(x) = 0$ otherwise. Then

(i) *x exists in A* $=_{df} (\chi_A(x) = 1)$;

(ii) *x does not exist in A* $=_{df} (\chi_A(x) = 0)$.

Remark 1 Note the condition that A be a well formed set, i.e. a set proper and not a meaningless array of symbols such as 'The set of all objects differing from themselves'. *Remark 2* It might be rejoined that Definition 3.27 could be replaced by a simpler one, namely

$$x \text{ exists in } A =_{df} x \in A.$$

True, but the membership relation is not a function, hence it does not allow one to take the next step, which is the introduction of an existence predicate:

DEFINITION 3.28 The *relative* (or *contextual*) *existence predicate* is the statement-valued function

$$E_A : A \rightarrow \text{Set of statements containing } E_A$$

such that $\ulcorner E_A(x) \urcorner$ is true if and only if $\chi_A(x) = 1$.

Therefore the old boring question whether existence is a predicate is seen to be ambiguous: the answer depends on whether $\exists$ or E_A is concerned. Whereas the particularizer or quantifier is not a predicate (statement valued function), the concept of relative existence we have just introduced, i.e. E_A, *is* a predicate.

The following examples show how to handle the concepts χ_A and E_A. In them 'M' stands for the set of characters in Greek mythology, 'c' for Chiron, 'C' for "is a Centaur", and 'W' for "is wise".

The centaur Chiron exists in Greek mythology.

$$Cc \;\&\; \chi_M(c) = 1, \quad \text{or} \quad Cc \;\&\; E_M c.$$

Some of the centaurs (existing) in Greek mythology are wise.

$$(\exists x)(Cx \;\&\; Wx \;\&\; \chi_M(x) = 1), \quad \text{or}$$
$$(\exists x)(Cx \;\&\; Wx \;\&\; E_M x).$$

All of the centaurs exist in Greek mythology and none of them is real.

$$(x)(Cx \Rightarrow (\chi_M(x) = 1 \;\&\; \chi_\Theta(x) = 0)), \quad \text{or}$$
$$(x)(Cx \Rightarrow (E_M(x) \;\&\; \neg E_\Theta(x))).$$

χ_A and E_A are concepts of unqualified relative existence as long as the reference set A remains unspecified. In tune with Postulate 3.4 in Sec. 1.3 we must distinguish two specific concepts of existence, namely those of conceptual existence (or existence in a conceptual context) and real existence (or existence in the world). The corresponding definitions are obvious:

DEFINITION 3.29

(i) *x exists conceptually* $=_{df}$ For some set C of constructs, $E_C x$;

(ii) *x exists really* $=_{df}$ For some set Θ of things, $E_\Theta x$.

For example the Pythagorean theorem exists in the sense that it belongs in Euclidean geometry. Surely it did not come into existence before someone in the Pythagorean school invented it. But it has been in conceptual existence, i.e. in geometry, ever since. Not that geometry has an autonomous existence, i.e. that it subsists independently of being thought about. It is just that we make the indispensable pretence that constructs exist provided they belong in some body of ideas – which is a roundabout fashion of saying that constructs exist as long as there are rational beings capable of thinking them up. Surely this mode of existence is neither ideal existence (or existence in the Realm of Ideas) nor real or physical existence. To invert Plato's cave metaphor we may say that ideas are but the shadows of things – and shadows, as is well known, have no autonomous existence. Nevertheless conceptual existence must be reckoned with not only to account for logic, mathematics and semantics, but also to better understand real existence.

Definition 3.29(ii) asserts that *all* things, and *only* things, possess the property of existing really – a property represented by E_Θ. This vindicates Aristotle's principle that *real existence is singular*. There are no general things: every real existent is an individual. (Therefore 'general systems theory' is a misnomer for 'general theory of systems'.) Whatever is general is either a property (e.g. a law) or an attribute (in which case it may be called a universal) or a proposition or a set of propositions (e.g. a theory).

A first consequence of Definition 3.29 is the non-Aristotelian principle that only fully qualitied objects are real. In particular neither the bare substance without properties nor the property without a substratum are real: they are so many fictions. Also: *pace* Quine, being the value of a bound variable does not guarantee existence – not even shadowy conceptual existence.

A second consequence of our identifying being with being a thing (or being a member of the set of things) is

THEOREM 3.7 (i) The world exists really. (ii) Every part of the world exists really.

Proof Part (i) follows from Postulate 3.3, Definition 3.4, and Definition 3.29. Part (ii) is entailed by Corollary 3.2 jointly with Definition 3.29.

Remark 1 Our postulated identity of existents and things amounts to stating that the world is *the world of things* and that no other world exists really. The extra worlds invented by the idealist philosophers as well as by Chwistek (1949) and Popper (1968) are so many fictions. *Remark 2* The identity of existents and things does not debase thoughts: it just denies the latter an independent existence. In Vol. 4, Ch. 10, we shall construe "person x thinks idea y" as "x brains y" much in the way as "x does y". *Remark 3* We are not identifying reality with actuality: for all we know reality might be the union of actuality and real possibility. More on this in Ch. 4, Sec. 2.4. *Remark 4* The existence function χ_Θ and the existence predicate E_Θ allow us to characterize very succinctly the main contenders over concrete or material existents:

Materialism $(\exists x)(\chi_\Theta(x)=1$, or $(\exists x)E_\Theta x$,

Immaterialism $(x)(\chi_\Theta(x)=0)$, or $(x)\neg E_\Theta x$.

4.4. *Nothingness and Virtual Existence*

To be, to exist really, is to be a thing. Consequently nonbeing, or nothingness, is failing to be a thing. There are several ways of handling the ontological concept of nothingness, depending on the context. These are some:

(i) b does not exist really: $\neg(b \in \Theta)$, or $\chi_\Theta(b)=0$, or $\neg E_\Theta b$;

(ii) there is nothing of kind K: $K=\varnothing$, or $(x)(\chi_K(x)=0)$, or $(x)\neg E_K x$;

(iii) there exists nothing with property P: $\mathscr{S}(P)=\varnothing$, or

$$(x)(\chi_{\mathscr{S}(P)}(x)=0), \text{ or } (x)\neg E_{\mathscr{S}(P)}x.$$

In each case nothingness or nonbeing is identified with nonexistence, not with some positive entity or even with some positive property. All three statements are negative, as they amount to either "There are no elements in the given set", or "The given individual does not belong to the given set". Therefore it is nonsense to regard nothingness as an entity, the way Heidegger and Sartre do, and even more to claim that existence is the synthesis of being and nothingness, as Hegel did.

The concepts of being (a thing) and nonbeing (or nothingness) are dichotomic. In other words, χ_A is a two valued function, and E_A is a unary predicate. That is, there are no intermediate degrees of existence, whether real or conceptual: the expression 'degree of existence' (or 'degree of reality') is a *flatus vocis*. Aquinas and some of his followers (e.g. Maritain) notwithstanding, there is nothing between being and nothingness: there are no degrees of being. For the same reason it is nonsense to hold that, while perceptible things may well exist, their atomic constituents have no autonomous existence but are either mere mathematical formulas or else an outcome of repeated human observations – as the Copenhagen interpretation of the quantum theory used to maintain.

However, there are two puzzling concepts in contemporary theoretical physics that, *prima facie*, denote neither existence nor nothingness. One is the concept of a vacuum occurring in quantum field theories, the other is the concept of a virtual quantum (e.g. particle, or photon). We shall presently show that their status is not the same: that the physical vacuum is a thing all right while a virtual particle is a fiction.

The vacuum that quantum field theories refer to is the field in its lowest stationary energy state. When the field is in this state it does not cease being a field: only, it has no quanta (photons in the electrodynamic case). Thus it is a hypothesis of quantum electrodynamics that, although there may be no electrically charged matter in a given region of space, there is always an electromagnetic field – the zero field or ever present fluctuating background. So much so that the vacuum is assigned definite physical properties, which become manifest when an intruder (matter or radiation) disturbs it. In sum, *vacuum field* $\neq$ *nothingness*. All that happens is that the word 'vacuum' has become inadequate, for it has acquired a new signification – it designates a new concept. The new picture of the physical world synthesizes atomism and plenism, by assuming that every cranny of the world is filled with things, although matter and radiation exist only in discrete units or quanta and not everywhere.

As for virtual particles and virtual photons, they are alleged to be extremely short lived and in principle unobservable, and to take part in (virtual) processes that do not conserve energy. For example, the interaction between a proton p and a neutron n is usually pictured as being embodied in, or transmitted by, a virtual pion π^+ emitted by p (i.e. $p \rightarrow n + \pi^+$) and then reabsorbed by n (i.e. $n + \pi^+ \rightarrow p$). The energy gained in the first virtual process is lost in the second, so that there is no net energy change. Besides, the period is so short (of the order of 10^{-24} sec) that the process is unobservable, whence the hypothesis is declared to be out of reach of experimental criticism. Strangely enough, the same people who accept these assumptions (*a*) recognize that the basic formulas of the theory entail the theorem of the conservation of energy, and (*b*) usually pay lip service to the operationalist philosophy.

Fortunately there is no need to take virtual particles seriously, i.e. to assign them a degree of reality midway between full reality and nothingness. The whole idea of a virtual particle stems from two mistaken philosophical assumptions. One is the naive realist belief that every term in an infinite series (or every branch in a Feynman diagram) must have a real counterpart. The other is the corpuscularian hypothesis that particles are the only realities, whence any interaction among them must be mediated by some of them – even if they have to be virtual. If both opinions are shed then the virtual particles and the shady transactions in which they are allegedly involved are seen to be idle fictions (Bunge, 1970a).

4.5. *Existence Criteria*

Our theory of things supplies no criterion for either establishing or refuting any hypothesis to the effect that such and such an object really exists. It is not the business of metaphysics to offer existence criteria, just as it is not within the province of the philosophy of mathematics to propose criteria for the existence of definite mathematical objects such as, e.g., the solutions to a given equation. However, a general indication can be given in either case.

In mathematics an object is said to exist in a given context if it satisfies certain conditions – e.g. an equation – i.e., if it has certain properties and holds certain relations to the other mathematical objects. Likewise in factual science something can be inferred to exist if it holds certain connections (not just relations) to something else whose existence has already been established or at least assumed. (Cf. Peirce, 1909, 6.318.) We sum up this idea in the following methodological

CRITERION 3.1 An object other than the entire world exists really if it is shown to be connected to some real object other than it.

This is a criterion of relative existence. Absolute existence cannot be established even though it need not be excluded. In order to show that x exists we must exhibit its connections with some thing y whose existence is not questioned in the given investigation. In particular, but not necessarily, this second object can be a human observer. In this case, if x is real then x will be a link in a chain ending up in an observer y, and the existence of x will make itself felt in some change in y – such as e.g. y's sudden perception of x.

Our reality criterion must not be taken for a *condition*, let alone a *definition*, of reality. Our definition of "reality" cannot be other than this:

DEFINITION 3.30 Let Θ be the set of all things and $[\Theta]$ its aggregation. Then

$$\textit{Reality} =_{df} [\Theta] = \Box = \textit{the world}.$$

The reality of an object consists in its being a part of the world. And the conjecture of the reality of an object must be *tested* through its immediate or remote perceptible effects but it does not *consist* in the latter. To say that it does condemns the greater part of reality to nonexistence and amounts to mistaking being for a being criterion – as operationism has been doing since the days of Peirce (cf. 5.406). If the world were to consist of a single indivisible physical thing, it would still exist or be real but there would be nobody to experience it and use or even reject Criterion 3.1.

Real existence statements should be taken in metaphysics just as seriously and as critically as they are taken in science. They must be regarded as pointing (rightly or wrongly) to actual or possible external referents, and also as being susceptible to correction in the light of fresh knowledge (i.e. new data or new conjectures or alternative theories). While methodical scepticism is indicated in this regard as in everything else, systematic scepticism won't do. That is, we may well doubt the existence of some particular thing but, if we want to check – i.e. to test the existence hypothesis – then we must, according to Criterion 3.1, assume without blinking, at least for the time being, the existence of something else. In practice we assume the existence of the entire world and we do not identify the latter with its explored part.

Assuming my own existence is necessary but insufficient for testing any existence hypothesis concerning some other thing. For one thing solipsism does not even account for my own birth, let alone for that of my enemy. Nor does solipsism explain the many disturbances – some agreeable, others disagreeable – impinging upon the subject. Not even phenomenalism suffices if we want to attain objectivity and efficiency. Consider the sentences 'There are cars' and 'Cars have at least three wheels'. To a realist, whether naive or critical, these sentences signify exactly and literally what they say. Not so to a phenomenalist: he will regard them as abbreviations of 'A subject has a car-like perception', and 'Whenever somebody has a car-like perception he can count at least three wheel-like perceptions' – or something of this sort. This subjectivism of his alienates him from science and can get him into trouble in everyday life. For, if the phenomenalist's car breaks down in the midst of a desert, when asking for a tow truck he might not say that *there is* a car stuck in the desert, let alone that there is nothing wrong with its wheels – that *they are* still on the car though unperceived. Unless he asks the Deity to do the perceiving for him in the meantime, the phenomenalist will have to contrive some complicated counterfactual sentences that no car mechanic would be interested in unravelling (e.g., 'If I were there now, I would perceive...'). For all practical purposes the phenomenalist will have to adopt a realist philosophy – unless he is willing to let his car go up in philosophical smoke. In short, the phenomenalist is bound to do double talk.

Finally, as for the concept of existence in a possible world and, even worse, that of existence in several different possible worlds, we shall leave them to the possible worlds metaphysician. The real world ontologist, busy as he is trying to figure out what the real world is like, has not time to waste with those figments of the idle imagination.

5. Concluding remarks

We have built a concept of thing out of the notions of a bare individual (Ch. 1) and of a property (Ch. 2). We have defined a thing as a substantial individual possessing substantial properties. And we have identified being, or real existence, with being a thing. Moreover we have assumed that the world contains only things. In fact we have defined reality as the aggregation of all things and indeed nondenumerably many of them.

There are conceptual objects to be sure but not as constituents of the world, much less constituting a realm of their own. Constructs, whether busy or idle, are fictions. Hence although in our ontology every class of objects is split into a class of things and one of constructs, the latter are not assigned an independent existence. It is not just that only individuals are real: only individual things – whether simple or aggregate – exist really.

Things can be represented schematically by what we have called functional schemata. A functional schema is a certain set equipped with a state function having a certain number of components. This state function is subject to restrictions – the law statements supposed to represent objective patterns. These laws limit the conceivable states a thing can be in: they restrict the domain of the state function for the thing. The set of all nomologically possible states of a thing constitutes its state space – or rather the state space for the thing in the given representation or functional schema.

Things come in natural kinds or species, i.e. classes of things possessing ("obeying") the same laws. The family of natural kinds or species does not have a Boolean structure but one of its own: it is a semilattice. This is due to the lack of isomorphism between the set of substantial properties and the set of predicates. A natural kind constitutes a natural grouping because it rests on a bunch of laws, but it is not a real thing: it is a construct – only, not an idle one.

The world consists in and of things but not every member of the class of things is actual. There are possible things, such as my great-great-grand children. Hence there are possible properties and possible facts. Besides, every actual thing has some actual properties and others which it may – or may not – acquire. In other words we are tacitly acknowledging real possibility. And we need this concept before we tackle the problem of change, as nothing will happen unless it is possible to begin with – i.e. unless there are things with possible states. Let us then turn to the problem of possibility.

CHAPTER 4

POSSIBILITY

We have assumed that the world is constituted by things only (Ch. 3). But things change and, as Aristotle saw, if a change occurs then it was possible to begin with (*Metaphysics* IX, 3). The seed germinates because it has the potentiality to do so. A becomes B only if it is in the nature of A that A *can* turn into B: this possibility of A's is one of its properties. (The converse is false: A may have the capacity to become B but this possibility may not be actualized due to adverse circumstances.)

Possibility, then, is inherent in reality because reality is changing. In other words, there are real possibles not just conceptual ones. An immutable world would be deprived of real possibility – and so would a changing world subject to inscrutable and unworldly Fate. But reality is neither unchanging nor chained to Fate, so there are real possibles. That is, we may partition reality into actuality and real possibility. (If on the other hand reality were identical with actuality, there could be no real possibility.) Therefore our metaphysics is possibilist not actualist.

Still, recognizing possibility, even objective possibility, does not entail regarding it as irreducible. Indeed an actualist may acknowledge possibility but will attempt to construe it in terms of actuality. He is bound to mistake real possibility for its test or for its measure. Thus he will tend to define the possible as that which either is or will be (Diodorus Chronos). Or he will assert that 'possible' means "sometimes" while 'necessary' means "always" (Russell, 1919). Or at least he will slight the importance of possibility by rendering it dependent upon actuality or somehow grounded in it – as is the case with Sellars (1963, Ch. 3) and Smart (1963, p. 23).

We shall see in this chapter that there is indeed a sense of 'possible' which is partially reducible to actuality, as when a child given the opportunity to eat a candy won't fail to actualize his natural bent or disposition to do so. This kind of possible property, that becomes actualized whenever the necessary conditions occur, will be called *causal disposition.* However there is another, deeper concept of possibility that is not so reducible and seems to have escaped the attention of most philosophers. An atom in an excited state can decay – with or without an external disturbance – into any of a number of lower energy levels none of them being predetermined. And the shuffling of genes that occurs in the fertilization of an egg cell is likewise random: any of a huge number of possible genomes can be realized. This kind of real possibility, to be called *chance propensity*, is not accounted for by actualism.

Real possibility will then be admitted as an ontological category not to be confused with either conceptual possibility or uncertainty. And we shall begin by distinguishing various kinds of possibility. First of all we shall study the concepts of conceptual possibility – alas, only to show that they are philosophically secondary because thoroughly reducible. Thereupon we shall analyze the two concepts of real or physical possibility. In the course of our study we shall have no occasion to employ modal logic, which will turn out to be too coarse a tool.

1. CONCEPTUAL POSSIBILITY

1.1. *Possibility Concepts*

There are several possibility species. They fall into either of two genera: *conceptual possibility* and *real possibility*. The former concerns formulas (in particular propositions) while the latter concerns factual items. Having such utterly different referents one may well wonder whether there is such a thing as a neutral concept of possibility subsuming those two, as modal logic presupposes. In any case, before analyzing the various concepts of possibility let us exhibit them globally: see Table 4.1.

1.2. *Four Concepts of Conceptual Possibility*

In the following 'K' will designate a body of knowledge – a set of data, conjectures, theories, etc. We propose first of all

DEFINITION 4.1 Let p be an arbitrary formula and A a subset of a body of knowledge K. Then

(i) p is *logically possible relative to* $A =_{df} A$ does not entail $\neg p$;

(ii) p is *mathematically possible relative to* $K =_{df}$ There is a model M in K such that p is satisfiable (formally true) in M;

(iii) p is *epistemically possible relative to* $K =_{df} p$ and K are mutually compatible (i.e. p contradicts no member of K);

TABLE 4.1
Concepts of possibility

Genus	Species	Referents	Meaning sketch
Conceptual or *de dicto*	Logical	Statements	Non contradiction
	Model theoretic	Partially interpreted formulas	Satisfiability in some model
	Epistemic	Factual statements	Consistency with what is known
	Methodological	Formulas	Provability or confirmability
Real or *de re*	φ-ontic	Physical facts	Lawfulness
	ψ-ontic	Psychical facts	Conceivability (compatibility with psychological laws and with background of subject)
	Deontic	Human actions	Not prohibited (by moral or positive law)
	Pragmatic	Human actions	Feasibility

(iv) p is *methodologically possible relative to* $K =_{df}$ There is no method m in K such that tests run with m disconfirm p relative to K.

(v) p is *conceptually possible relative to* $K =_{df}$ p is either logically or mathematically or epistemically or methodologically possible relative to K.

Example 1 Any noncontradictory proposition is logically possible, for all systems of logic admit the principle of noncontradiction. *Example 2* The formula "$x^2 = -1$" is mathematically possible since it is satisfied e.g. in the field of complex numbers (namely by $x := i$). *Example 3* For all we know Oparin's theory of the origin of life might be true. *Example 4* The predictions calculated with the help of specific scientific theories are methodologically possible since their negates are refutable in principle.

Remark 1 Our definition reduces conceptual possibility to conceptual actuality. Consequently it renders the concept of conceptual possibility redundant – in principle though not in practice. *Remark 2* The various concepts refined by Definition 4.1 are concepts of *conditional or relative possibility*. Absolute possibility can be construed as hidden conditional possibility, by just adding suitable existential quantifiers. For example, on the basis of Definition 4.1(i) we could make the following definition:

p is logically possible $=_{df}$ There is no set of premises entailing $\neg p$. But one wonders what the use of this concept could be, as every responsible judgement is made by taking some definite premises for granted. *Remark 3* Our concepts of conceptual possibility satisfy no system of modal logic. In particular logical possibility, as defined above, satisfies only one of the two axioms common to all modal logics, namely "If p then it is possible that p". (See Hughes and Cresswell, 1968.) Besides, the standard systems of modal logic deal with absolute not conditional (relative) possibility.

The only ingredient of modal logic we do need is Aristotle's definition of necessity ($\Box$) in terms of possibility ($\Diamond$), namely:

p is necessary $=_{df}$ It is not possible that $\neg p$, or $\Box p =_{df} \neg\Diamond\neg p$. This definition, together with Definition 4.1, entails

DEFINITION 4.2 Let p be an arbitrary formula and A a subset of K. Then

(i) p is *logically necessary relative to* $A =_{df} A$ entails p;

(ii) p is *mathematically necessary relative* to $K =_{df} p$ is satisfied in every model contained in K;

(iii) p is *epistemically necessary relative to* $K =_{df} K$ implies p.

(iv) p is *methodologically necessary relative to* $K =_{df}$ For all methods m in K, tests on p run with m confirm p relative to K.

(v) p is *conceptually necessary* relative to $K =_{df} p$ is either logically or mathematically or epistemologically or methodologically necessary relative to K.

Again, none of these concepts of conceptual necessity coincides with those handled by modal logic. To begin with, they are concepts of relative (contextual) not absolute (context free) necessity. Besides, our notion of logical necessity, which reduces to the entailment relation, violates the modal axiom: $\Box(p \Rightarrow q) \Rightarrow (\Box p \Rightarrow \Box q)$. Instead it satisfies $\Box q \Rightarrow \Box(p \Rightarrow q)$, as can be seen with the help of the principle of conditional proof. (Anyway, what would be gained by constructing a deviant modal calculus?) In sum we have no use for modal logic because it is incapable of elucidating any useful notion of conceptual possibility and – to anticipate a result to be obtained in the next section – because it is equally impotent in tackling the notion of real or physical possibility.

Though for different reasons, we concur with the verdict that "Modal logic is of no philosophical significance whatsoever" (Bergmann, 1960).

1.3. *Conceptual Possibility*: *Relative*

We close by emphasizing the relativity of conceptual possibility as well as its difference from factual or real possibility. Formulas, just like actual facts, are neither possible nor impossible in themselves: they just are. What is possible or impossible with regard to a formula is some relational property of it such as exemplifying it, or proving it, or confirming it, or showing that it is compatible with a certain body of knowledge. In other words possibility is not inherent in formulas but is a relation between them and certain other conceptual items. That is, a statement of the form "It is possible that p" is not on the same level as p itself but is a metastatement. Moreover it belongs in pragmatics not in semantics, let alone in ontology.

Not so states of affairs: these are either actual or possible absolutely, in the sense that their possibility does not depend upon any body of knowledge. Propositions such as "That atom may disintegrate within the next minute", and "This organism is viable", are object statements not metastatements; they involve the notion of real possibility not that of conceptual possibility. Moreover these two concepts are unrelated except linguistically. To begin with they have nonoverlapping reference classes: while the referents of the concept of conceptual possibility are propositions, those of the concept of real possibility are facts. Hence their meanings are radically different. Therefore the attempt to have a single theory (namely modal logic) cover both concepts is bound to fail just as miserably as the attempt to describe wedding rings and algebraic rings with the same theory.

So much for conceptual possibility.

2. REAL POSSIBILITY

2.1. *Fact*

Unlike the concept of conceptual possibility, that of real (or physical or ontic) possibility refers to factual items. That is, the statement that p is really possible assigns the factual referent(s) of the proposition p, not p itself, the property of being really possible. So much so that a proposition of the form "p is really possible", where p describes some fact, may in turn be conceptually impossible relative to some body of knowledge.

Our referents are then factual items: things, properties of things, kinds of thing, states of things, and changes of state of things. Factual items are represented by definite descriptions and propositions such as the following:

Thing b	Thing b exists.
Property P	Thing b possesses property P.
Kind K	Thing b is of kind K.
State s	Thing b is in state s.
Event e	Thing b undergoes change e.

The concepts occurring in the above, with the exception of that of event, have been elucidated in the previous chapter; the notion of change will be clarified in the next chapter. We proceed to single out two kinds of factual items, namely being in a given state and going from one state to another:

DEFINITION 4.3 Let X be a thing. Then f is a *fact* involving X iff either

(i) f is a *state* of X, i.e. there is a state space $S_{\mathbb{L}}(X)$ for X such that $f = s \in S_{\mathbb{L}}(X)$, or

(ii) f is a *change of state* of (or *event* in) X, i.e. there is an $S_{\mathbb{L}}(X)$ such that $f = e = \langle s, s' \rangle \in S_{\mathbb{L}}(X) \times S_{\mathbb{L}}(X)$.

Note that, as here conceived, a state is always a state of some thing, and likewise an event is a change in the state of some thing. Note also that the states in question are only those which are lawful, as indicated by the subscript $\mathbb{L}$. (However, there is an ambiguity with regard to 'event', since although a pair $\langle s, s' \rangle$ may be lawful it may well happen that there is more than one way to go from s to s'. In other words, $\langle s, s' \rangle$ does not represent a single event but rather a whole collection of events. This ambiguity, to be removed in Ch. 5, won't do any harm at present.)

Let us now turn to a structural characterization of the notion of a fact. Call F the set of facts as defined by Definition 4.3. This set is of course the domain on which the predicates "is really possible" and "is actual" (or "is the case") are defined. (For example, we have "State s of thing b is really possible" and "Thing b is actually in state s".) These two concepts can be characterized in simple terms, namely thus. Assume, as is usual in the applications of probability theory, that every single factual item is construed as (represented by) a set – even if the latter is just a singleton or even void. That is, the set F of facts will be regarded as a family of subsets of some set – never mind which one. Stipulate that, if x is in F and is nonvoid, then x is an unspecified possible fact, whereas if $x = \varnothing$ then x is impossible. Assume further that possibles can conjoin, disjoin and invert. That is, if x and y are really possible factual items (i.e. nonempty members of F) so are $x \cap y$, $x \cup y$, and $x^{-1} = F - x$. The intersection of x and y represents the compossibility of x and y, the union represents the alternative possibility, and the inverse or complement x^{-1} of x represents whatever is possible when x is not. (Note that x^{-1} is not a fact but a set of facts.)

On the other hand actuals either occur or they don't, and some of them occur jointly but never disjointly. Indeed there is no such thing as the actual fact that fact *a or* fact *b* happens, even though the corresponding proposition may be true. Nor does x^{-1} happen. There is no such thing as the unique complement or inverse of an actual fact. Complementation and disjunction are marks of possibility (as well as of constructs) not of actuality. Further, if x and y are really possible, we call $A(x)$ the actualization or occurrence of x, $A(x^{-1})$ the nonoccurrence of x, $A(x \cap y)$ the occurrence of both x and y, and $A(x \cup y)$ that of at least one of the facts. (See Table 4.2.) And we lay down

POSTULATE 4.1 Let F be a collection of sets, $\cap$, $\cup$, and $^{-1}$ the boolean operations on F, and A a function from F into a set of propositions. Then the structure $\mathscr{F} = \langle F, \cap, \cup, ^{-1}, A\rangle$ is a *space of facts* iff

(i) $\langle F, \cap, \cup, ^{-1}\rangle$ is an algebra of sets;

(ii) F represents the set of facts (actual or possible) involving some thing or things X;

(iii) A is a function from F into the set of propositions concerning the thing or things X, such that

TABLE 4.2

Real possibility and actuality. Proviso: $x, y \neq \varnothing$

Formula	Interpretation	Formula	Interpretation
$x \in F$	x is a possible fact	$A(x)$	x occurs
$x \cap y \in F$	x and y are compossible	$A(x \cap y)$	x and y occur
$x \cup y \in F$	x or y is a possible fact	$A(x \cup y)$	x occurs or y does
$x^{-1} \subset F$	that which is possible when x is not, is a possible fact	$A(x^{-1})$	x does not occur

(a) for any $x \in F$, $A(x^{-1}) = \neg A(x)$;

(b) for any $x, y \in F$, $A(x \cap y) = A(x) \,\&\, A(y)$ – i.e. A is a morphism of opposition and multiplication;

(iv) for any $x, y \in F$ other than the empty set, $A(x)$ represents the occurrence (actualization, realization) of (the possibility) x, $\neg A(x)$ the nonoccurrence of x, and $A(x) \,\&\, A(y)$ the occurrence of both x and y.

It is easily shown that the actualization function A has the following further properties:

THEOREM 4.1 Let A be the propositional function on F defined by Postulate 4.1. Then for all $x, y \in F$

(i) $A(x \cup y) = A(x) \vee A(y)$;

(ii) $A(x^{-1} \cup y) = A(x) \Rightarrow A(y)$;

(iii) $A(x \cap x^{-1}) = A(x) \,\&\, \neg A(x)$.

The last result is of course the reason for interpreting $\varnothing$ as impossibility.

Our Postulate 4.1 is at variance with an alternative set of axioms of occurrence (Suppes, 1970, p. 38), supposed to characterize the notion of occurrence or actualization. These axioms do not discharge this function if only because the first of them states that, if x and y are events, and x occurs, then their union $x \cup y$ occurs as well. In our ontology there are no disjunctive facts: there are only disjunctive possibilities and disjunctive propositions about facts whether possible or actual.

Postulate 4.1 constitutes only a partial characterization of the possibility/actuality contrast. It tells us just that (*a*) the set of occurrences is a proper subset of the set of possibilities, and (*b*) the algebraic structure of the latter is richer than that of the former. The actualization process is somehow representable as a collapse of the richer structure into the poorer one: in this process union (or disjunction) and complementation (or negation) are forgotten. However, this characterization in terms of a forgetful functor is insufficient because it fails to specify (*a*) which subset of the conceptually possible facts (for this is what F is) constitutes the *really* possible facts and (*b*) which subset of the collection of really possible facts gets actualized. These defects cannot be repaired by proposing a more complex theory in the same vein: no a priori theory, independent of considerations concerning laws and circumstances, can single out the set of really possible facts, let alone that of actuals. Let us then get in touch with reality.

2.2. *Chrysippian Possibility*

The profound Stoic philosopher Chrysippus defined the possible as "that which is not prevented by anything from happening even if it does not happen". Contrast this view to that of Diodorus Chronos, according to whom "the possible is that which either is or will be true". (Cf. Kneale and Kneale, 1962; Rescher and Urquhart, 1971.) According to this other view the concepts of actuality and possibility are coextensive: that which never happens is impossible. (See also Hartmann, 1938.) Consequently the notions of unrealized potentiality and of missed opportunity find no room in this conception. Yet, as Aristotle and Chrysippus saw, not all possibles are realized: "thus it is possible that this gem will break even if it never does" (Cicero, *De fato* IX).

Chrysippus' definition of real possibility hinges on the notion of lack of constraint or absence of inhibition. The latter notion can be construed as follows:

DEFINITION 4.4 If x and y are possible facts (i.e. $x, y \in F$) then x ***inhibits*** (or ***prevents***) the occurrence of $y =_{df} A(x) \Rightarrow \neg A(y)$.

Chrysippus' idea is then perhaps rendered by

DEFINITION 4.5 A fact is *really possible* iff there is no actual fact the occurrence of which prevents it: If x is a fact then

$$\Diamond_r x =_{df} \neg(\exists y)(y \text{ is a fact } \& \ y \text{ occurs } \& \ y \text{ prevents } x)$$

$$\Leftrightarrow \neg(\exists y)(y \in F \ \& \ A(y) \ \& \ (A(y) \Rightarrow \neg A(x))).$$

In other words, anything is (really) possible that is unhampered or free to happen. Freedom, whether physical or moral, is thus equated with real possibility. (Physics and ethics acquire therefore a common metaphysical basis.) The concept of real impossibility is obtained by negating the two sides of the previous definition: a fact is *really impossible* iff it is inhibited by some other fact. We infer then that if anything is impossible then it is not the case. This corollary is just the contraposition of "Whatever is the case is possible".

We shall not elaborate on the Chrysippian notion of possibility because we shall adopt an alternative conception of possibility, namely the nomological one to be expounded presently.

2.3. *Real Possibility as Lawfulness*

Our point of departure will be Bolzano's conception of possibility as lawfulness: "physically possible is that which does not contradict any of the so called laws of nature" (Bolzano, 1821, art. 'Möglich', p. 65). However, we cannot adopt this definition literally because we regard laws as objective patterns not as propositions and consequently cannot make use of the concept of consistency occurring in Bolzano's statement.

To begin with recall the Definition 4.3 of a fact as either a lawful state or a lawful change of state of a thing. We shall assume that these are the only really possible facts. That is, we identify real possibility with nomic possibility. More explicitly, we assume

POSTULATE 4.2 A fact f is *really possible* iff f is a lawful fact, i.e. iff either $f \in S_{\mathbb{L}}(X)$ or $f \in S_{\mathbb{L}}(X) \times S_{\mathbb{L}}(X)$ for some thing X and some state space $S_{\mathbb{L}}(X)$. In symbols,

$$\Diamond_r f \text{ if and only if } f \text{ is a lawful fact.}$$

Note that this is not a definition but a hypothesis and moreover one used surreptitiously when framing Definition 4.3. In fact Postulate 4.2 can be ignored or even negated. An empiricist ontology – one recognizing no objective laws, as was the case with Hume's and James' – finds no use for that assumption. And a religious metaphysics, unless it be of the Cartesian or Leibnizian kind where the Deity is always strictly law-abiding, will negate our axiom.

The following consequences follow immediately from our postulate:

(i) *Lawless individual facts are impossible.*

(ii) If, with Hume, we define the miraculous as the lawless, then we infer that *miracles are impossible.* If this conclusion is rejected then the concept of real possibility must be redefined – or just dropped, perhaps together with that of an objective law or pattern. Either alternative would be consistent with an empiricist ontology.

(iii) Randomness too has been defined as lawlessness, but this is not satisfying. First, whereas the predicate "lawful" applies to individual facts, "random" applies only to sets of facts – such as e.g. a sequence of hits and misses of shots aimed at an object with unknown location. Second, randomness is a kind of lawfulness not its opposite. What is the opposite of lawfulness is chaos or the absence of all laws, even stochastic (probabilistic) laws. But chaos too is a property of sets of facts not of individual facts. For example, by collecting haphazardly a number of facts (possible or actual) concerning unrelated things we can form a lawless collection of individually lawful facts. (Think of a set made up of an arbitrary collection of astrophysical, genetic, and political events.) In sum randomness – a basic feature of reality if we are to believe contemporary science – is, unlike chaos, compatible with lawfulness or nomic necessity. More on these matters in Sec. 6.4.

(iv) We conclude the falsity of the identification of lawfulness and necessity, popular since the 17th century, made obsolete by the birth of statistical theories from the 1870's on, assailed by Peirce, and recently revived by Montague (1960, repr. in 1974). That identity is in fact a cornerstone of Montague's possible-world semantics and metaphysics – of which more in Sec. 6. In short lawfulness is not identical with necessity but with (real) possibility. More on necessity anon.

2.4. *Factual Necessity*

Conceptual necessity, it will be recalled, can be introduced in terms of conceptual possibility by way of Aristotle's definition, namely: $\Box p = \neg\Diamond\neg p$. We cannot use this definition to obtain the notion of nomic necessity out of that of nomic possibility as coextensive (though not cointensive) with that of lawfulness. For one thing negation applies to propositions, not to facts: the expression '$\neg x$' makes no sense when x denotes a fact. We might of course reinterpret '$\neg x$' as "It is not the case that x", i.e. as $\neg A(x)$. But even so Aristotle's definition would not help us, because real or factual necessity has a component absent from lawfulness, namely *circumstance.*

Indeed for anything to actually happen and thus be (really) necessary it must not just "obey" certain laws but also "count on favorable circumstances". Not even a deterministic (nonstochastic) law statement describes only what is actually the case: *every law statement describes possibles* – without of course the help of modal operators. (Think of the bundle of trajectories in the state space representing the set of solutions of an equation of evolution.) A fortiori, stochastic laws describe even weaker possibles, namely chance facts, i.e. facts occurring only with a fixed frequency. In sum we may keep Aristotle's definition of possibility for the conceptual realm – as we did in Sec. 1.2 – but it is inapplicable in science and in scientific ontology. Here we need rather

DEFINITION 4.6 Let $x \in F$ be a really possible (i.e. lawful) fact. Then

(i) x is *necessary* iff there is another fact $y \in F$, called the *circumstance attending* x, such that $A(y) \Rightarrow A(x)$;

(ii) x is *contingent* iff x is not necessary.

For this reason actual states of affairs, or situations, are best accounted for (described, explained or predicted) with the help of both law statements and propositions representing the circumstances (idiosyncrasies, initial conditions, boundary conditions, etc.) jointly necessary for the situations of interest to actually occur either always or with a fixed frequency. In other words, the preceding considerations constitute the ontological ground for the familiar schema of nomological explanation:

$$\text{Laws, Circumstances} \vdash \text{Facts.}$$

At first sight the necessity or contingency of a fact or situation depends exclusively on the kind of law involved: deterministic laws would cover necessary facts, nondeterministic laws contingent facts. For better or worse this simple schema is false. In fact deterministic laws can cover contingent facts and stochastic laws can cover necessary facts. A familiar example of the former kind is the impact of a classical particle on a wedge: see Figure 4.1. In this case there are two equally probable trajectories, i.e. the problem has two different solutions on an equal footing. And a familiar example of the second type is the repeated

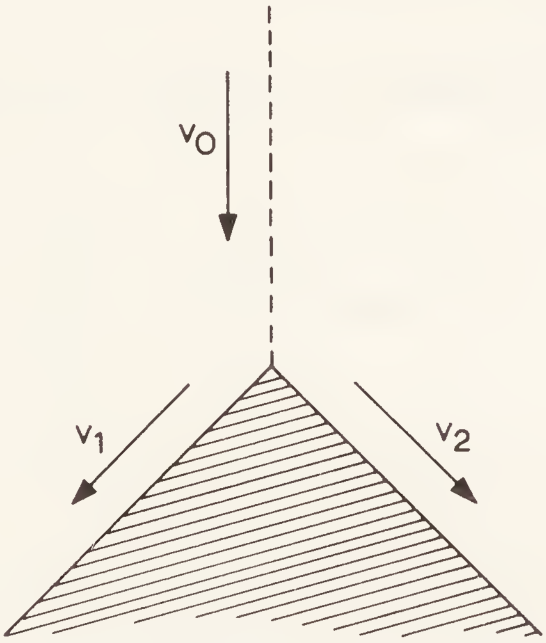


Fig. 4.1. A particle (or a fluid jet) collides against a rigid wedge and is deflected either left or right with the same speed. For a classical discussion see Truesdell (1974).

tossing of a penny: in the long run certain regular sequences, such as five heads in a row, are bound to occur and in fact they occur with a fixed probability.

In sum, the explanatory schema $\ulcorner L, C \vdash \mathit{Fact} \urcorner$ lumps together four accounts:

Deterministic law, C_1
$\vdash$ Necessary fact (occurs whenever C_1 or C_2)

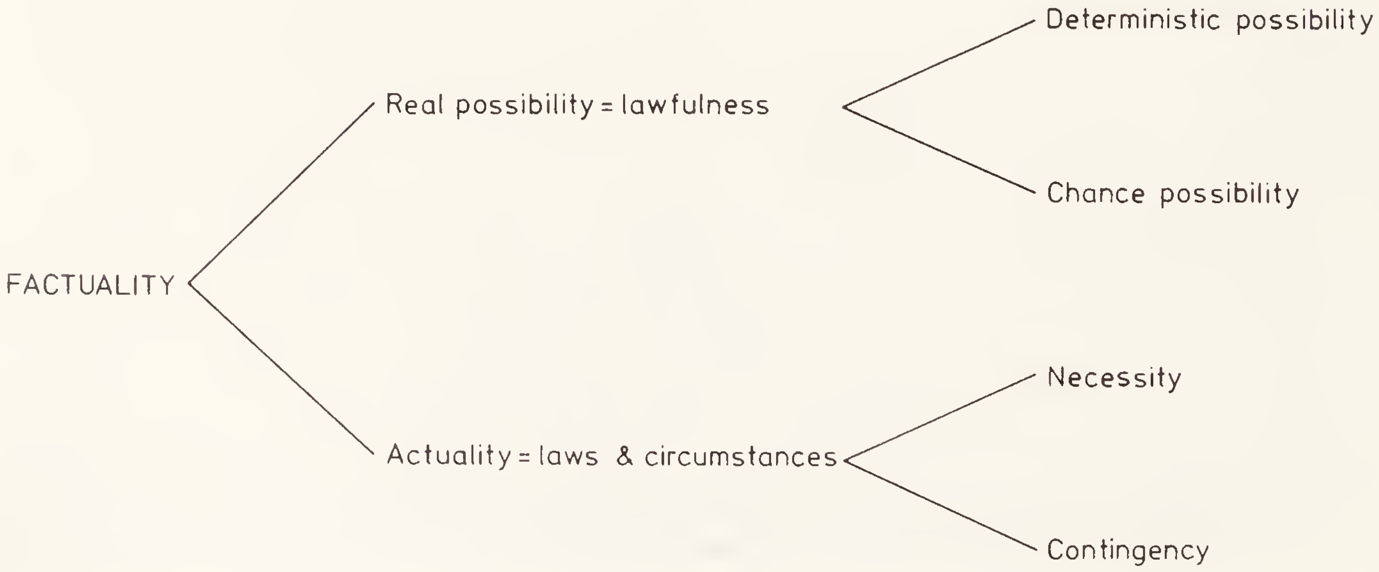
Stochastic law, C_2

Stochastic law, C_3
Contingent fact (occurs with fixed probability whenever C_3 or C_4)

Deterministic law, C_4

To conclude this subsection. Real possibility is identical with lawfulness. But real necessity is not definable in terms of possibility and negation. Never mind, for the concept of necessity plays hardly any role in factual science except as a synonym for actuality: whatever happens to be the case must be, and conversely. Little is gained by replacing "It is snowing" with "Necessarily it is snowing", or "The snowfall probability is $\frac{3}{4}$" with "Necessarily the snowfall probability is $\frac{3}{4}$", let alone with "The probability that it snows is necessarily $\frac{3}{4}$". *The prefix 'necessarily' is unnecessary*. An actual particular fact is a fact is a fact. Calling it 'necessary' is just a way of saying that, far from being merely possible, it is the case.

In sum, we have the following partition of the factual realm:


FACTUALITY
Real possibility = lawfulness
Deterministic possibility
Chance possibility
Actuality = laws & circumstances
Necessity
Contingency

2.5. *Possibility Criteria*

If scientific research is conceived of as being ultimately the search for objective laws (Bunge, 1967a, Vol. I, Ch. 6), and if lawfulness is equated with nomic possibility (Postulate 4.2), then science must be regarded as the study of the really possible. In fact every scientific activity of the theoretical kind deals not only with actual facts but also with real possibilities. Thus theoretical mechanics – unlike experimental mechanics and applied mechanics – studies all possible motions of all possible bodies. Theoretical chemistry – unlike experimental chemistry and chemical engineering – studies in principle all possible chemical compounds. And theoretical genetics studies all possible gene recombinations and mutations.

In other words all theoretical science utilizes the concept of a possible thing and even that of a possible state of a possible thing, much to the dismay of actualist philosophers. But of course the notion of a possible individual of this kind is not characterized with the help of either modal logic or model theory – let alone with that of any system of possible worlds metaphysics. In theoretical factual science a possible thing, such as a possible compound or a possible organism or a possible community, is an arbitrary member of the class of hypothetical referents of the theory of interest – i.e. an individual possessing ("obeying", "satisfying") the law statements of the theory. On the other hand the notion of a possible world (nowadays often identified with that of a model of an abstract theory) is alien to factual science: we have access to no substantial world other than the real world. True, possibility is inherent in the world, but there is no special carrier of possibility – no possible substantial world in addition to the actual world. Every concrete thing possesses certain possibilities to the exclusion of others.

What holds for theoretical possibility holds also, *mutatis mutandis*, for theoretical impossibility. Every theory rules out a number of possibilities just as it countenances others. Of course any theory can go wrong. But as the number and variety of theories agreeing on the impossibility of something increases, the likelihood of a "prohibition" increases. Even so one should keep an open mind. For example relativistic mechanics rules out the possibility that there be superluminal particles (tachyons). But the theory has nothing to say about noncorpuscular entities travelling faster than light. Moreover a good theory about such entities would help in searching for them.

Any theory hypothesizing the existence of a thing or a fact not envisaged by antecedent theories should respect the latter in their own "domains of validity" (extensions) and should enjoy the support of independent evidence or else stimulate the search for it. And any theory hypothesizing the existence of things "forbidden" by well corroborated theories should say whatever those theories tell us about the things those theories allow. Needless to say, even logic must be taken into account when listing impossibles. Metaphysics, on the other hand, is hardly in a position to admit or rule out any fact. What metaphysics can do is to clarify some of the concepts involved in scientific judgments of possibility or impossibility.

When shifting our attention from reality to science, talk of facts becomes talk of the knowledge of facts and, in particular, a definition of possibility may become a possibility criterion. If we were to adopt Chrysippus' concept of real possibility (Sec. 2.2) we would have to adopt the following criterion: *If x is a fact, then x is known to be really possible iff nothing is known to prevent x.* But since we have proposed a modification of Bolzano's notion of possibility (Sec. 2.3) we cannot accept the Chrysippian possibility criterion. We shall instead suggest a criterion inspired in the following example.

A theoretical chemist will declare an unknown compound to be really (chemically) possible iff (*a*) its components exist, (*b*) they can assume a stable configuration, and (*c*) a reaction mechanism resulting in the hypothesized molecule is theoretically possible, i.e. consistent with what we know. In short, the chemist's possibility statement is the conclusion of a (rather complex) argument involving items both theoretical and empirical. We generalize this procedure by proposing

CRITERION 4.1 Let T be a theory, and E a body of empirical evidence couched in the language of T and relevant to T, and suppose that both T and E refer to a fact x described by a proposition $p[x]$. Then x is *really possible according to* T *and* E iff $T \cup E$ fails to entail $\neg p[x]$.

This is a theoretical possibility criterion and it refers us back to what is already known (or presumed to be known), not to new observations or future theories. If a theoretically possible fact occurs it may count as a confirmer of the theory. If it fails to occur the theory remains unaffected and we have no case against the possibility of the fact. And if a theoretically impossible fact does occur then we shall have to revise either the theory or the evidence employed in inferring the impossibility.

Consider now the possibility criterion suggested by the Diodorean concept of possibility, namely: *Fact x is known to be possible iff x is observed to happen at some time.* This criterion is quite useless, as it works only *ex post facto*: it does not allow me to regard my own demise as possible until it is too late for doing any thinking of this kind. And once the fact is observed to happen it may as well be pronounced necessary. But if it is not observed to happen then the reasonable attitude is to wait and see rather than to declare it impossible. These are not defects that can be remedied while retaining the spirit of Diodorus' conception of possibility: they are inherent in any attempt to frame an *empirical* possibility criterion. Indeed such a criterion must refer to actuals and is therefore impotent in distinguishing possibility from actuality. In other words, the observation that something is the case only shows that it was indeed possible to begin with but does not suggest whether it was just possible or necessary. In conclusion, empirical criteria of possibility are impossible. We must remain content with a theoretical criterion like the above. Which is fitting, since the concept of possibility is theoretical not empirical.

3. Disposition

3.1. *Intuitive Idea*

Thus far we have been concerned with the general concept of real possibility. We shall study henceforth two special concepts of real possibility. One is *disposition*, or *causal propensity*, as exemplified by fragility and inherited susceptibility to TB. The other concept is that of *chance propensity*, as exemplified by a photon that may go through either slit of a two slit screen. Let us start with the former, which is the more familiar of the two – therefore probably not the more basic.

A force may cause a change in relative position, a lump of sugar may dissolve, a muscle cell may divide, a literate person may read, a viable society can survive. These potentialities will be realized provided the proper environment or means is supplied: otherwise they will not. (To repeat: Such potentialities are realized *whenever* suitable conditions are present.)

Moreover, as the Philosopher saw, potency precedes act. (And conversely, as is equally obvious. Thus, being born with a normal brain and being adequately fed are necessary conditions for a person to be capable of learning and doing certain things, even if he never actually learns or does them.) The disposition to do x is then prior to doing x. Thus a piece of matter will actually refract light provided it was refractive to begin with; and if it was refractive then it was so whether or not it was exposed to light. (Do not mistake a property for the tests for it.) Recall Newton's *Opticks* (1782 ed., Vol. IV, p. 6): "Refrangibility of the Rays of Light, is their disposition to be refracted or turned out of their way, in passing out of one transparent Body or Medium, into another". A light ray has all the time the refrangibility property, even while it propagates in a vacuum. Likewise a crystal retains its refrangibility while locked up in a chest.

A peculiarity of dispositions (or causal propensities) is that they come in pairs. Thus (all) light rays are refrangible iff (some) bodies are refrangible. And every lock can be opened by some key(s). Let us keep this complementarity in mind, for it will occur in the very definition of a dispositional property.

The peculiarity of a *causal* propensity, in contrast to a chance propensity, is that the former never fails to be actualized when conjoined with suitable circumstances. Solubility becomes actual dissolution, divisibility division, viability continued existence, and so on. In clumsy but few words: *Disposition & Circumstance = Actuality*. Equivalently: Disposition equals actuality minus certain circumstances. More precisely: A disposition is a condition that, though necessary, is not sufficient. Accordingly, if x is a fact, then:

$$\text{If } (\exists y)(y \in F \;\&\; y \neq x \;\&\; (A(y) \Rightarrow A(x)) \quad \text{then} \quad \Diamond_r x.$$

(The converse is false, as real possibility is not exhausted by causal propensity.) Real possibility of this kind, i.e. disposition, is then an insufficient condition: actuality ensues as soon as this shortcoming is remedied, i.e. when the missing conditions appear. (For a similar independent construal see Raggio, 1969.)

The circumstances favorable or unfavorable to the actualization of a potentiality of a thing x involve some thing y other than x and forming part of the environment of x. This other thing, to be called the *complement* of the first, must have a disposition matching that of the thing of interest, for the first disposition to actualize. Thus a certain key may open provided it is joined to a proper lock with the unlocking disposition. What exhibits an actual or manifest property is the whole formed

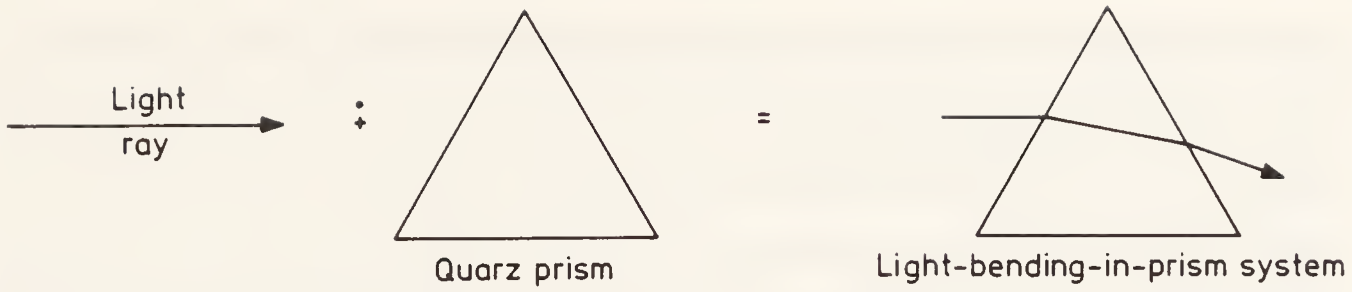


Fig. 4.2. Two entities, each with a certain disposition, aggregate into a third entity with a certain manifest property.

by the thing of interest and its complement. (See Figure 4.2.) The general schema is as follows:

> *Thing x with disposition P joins thing y with disposition Q to form thing* $z = x \dotplus y$ *with manifest property R.*

A few more examples should do no harm. *Example 1* If a piece of magnetizable material is placed in a magnetic field, it becomes magnetic. *Example 2* A fertilized bird's egg lying long enough in a nest at a suitable temperature becomes a chick. *Example 3* The inherited susceptibilities to TB, schizophrenia and other illnesses are parallel: the environment brings about, or else inhibits, such an inborn disposition.

We close this subsection by noting that the degree of a predicate is no indication of whether the property it represents is manifest or dispositional and, if the latter, whether it is causal or stochastic. Thus human abilities are often intrinsic properties (though of course displayed only under certain circumstances). On the other hand the velocity of a macrosystem is a mutual property (of thing and reference system) and a manifest one. Likewise with chance propensities. Thus the localizability of a quantum mechanical system is intrinsic in the case of a single "particle" and mutual for a pair of correlated "particles".

3.2. *Elucidation*

The preceding informal considerations are summed up and exactified in the following convention, which employs only concepts that have been clarified in the previous chapters:

DEFINITION 4.7 Let $z = x \dotplus y$ be a thing composed of different things x and y. Further, let P be a property of x and Q a property of y, with Q possibly equal to P. Then P is said to be a *disposition* (or *causal* *propensity*) of x, and Q the *complement* of P with respect to y, iff there is a third property R such that

(i) $z = x \dotplus y$ possesses R;
(ii) $z = x \dotplus y$ possesses neither P nor Q;
(iii) neither x nor y possesses R.

Note that what is called a *disposition* and what its complement is a matter of viewpoint or convention. Thus one says that sugar is soluble (in water), but this is because water has the disposition to dissolve sugar.

Example An aggregate of reactants won't start reacting unless the required pressure and temperature conditions become available – which conditions are properties of the immediate environment of the aggregate. When they do, the disposition to react in a certain way is actualized, and eventually new things (the reaction products) emerge that do not have all the properties of the reactants but are characterized instead by new properties determined by the former.

We are now in a position to elucidate a notion that has become incorporated into common knowledge:

DEFINITION 4.8 The totality of dispositions (causal propensities) of a thing is called the *causal potentiality* of the thing. I.e., if x is a thing then

$$\Pi_c(x) = \{P \in p(x) \mid P \text{ is a disposition of } x\}.$$

DEFINITION 4.9 One thing has a *greater causal potentiality than* another iff the potentiality of the former includes that of the latter. I.e., if x and y are things, then

$$\Pi_c(x) > \Pi_c(y) =_{df} \Pi_c(x) \supseteq \Pi_c(y).$$

Example 1 The potentiality of an individual organism consists of its genetic make-up or program. A poor genetic endowment yields few possibilities of a rich experience, just as a poor environment cuts out many possibilities of using the genetic make-up. *Example 2* Shortly before the birth of molecular biology it was considered preposterous to regard the human zygote as a potential man, or the acorn as a potential oak tree. Both assertions would have been laughed away as remnants of preformationism. But contemporary biology is preformationist insofar as it holds the whole mature organism to be programmed in the zygote. *Example 3* The psychological faculties or abilities, rejected by behaviorism, are back. Thus N. Chomsky, G. A. Miller and other psycholinguists stress the difference between ability and performance – an instance of the potentiality-actuality difference. Thus verbal behavior does not coincide with linguistic ability although it is the only overt hence observable manifestation of it. The ability is in this case the capacity to utter and understand sentences new to the speaker. But this potentiality can be either actualized or frustrated by the environment.

3.3. *Potency and Act*

To Aristotle potency was prior to act: actuality was the unfolding of potentiality. The latter was in turn left unexplained. The distinction between potency or disposition, on the one hand, and act or manifest property, on the other, has been kept by modern science – a fact suppressed by the positivist philosophy of science. Thus Newton: "Colours in the Object are nothing but a Disposition to reflect this or that sort of Rays more copiously than the rest" (Newton, 1782, Vol. IV). Other physical properties can be described with paraphrases of the latter statement: just think of conductivity, magnetic permeability, refrangibility, or viscosity. They are all powers or dispositions of the causal kind. Such examples have suggested the conjecture that "all physical (and psychological) properties are dispositional" (Popper, 1957, p. 70). But this would be impossible if only because the very concept of potency makes sense only relative to that of act – as shown e.g. by Definition 4.7.

Actualists have tried to ignore the long list of dispositions exhibited by modern science. They often argue against objective dispositions on the strength of an empiricist epistemology. They claim, and rightly so, that only actual things and manifest properties are publicly observable. (But they do not seem to realize that we cannot possibly do without dispositionals such as "observable".) They will try hard to define, and thus eliminate, dispositions in terms of observable properties. A favorite gambit of theirs is Carnap's method of introducing dispositionals by way of bilateral reduction sentences (Carnap, 1936–37), namely thus:

> Thing x subjected to test condition C is assigned disposition D just in case x exhibits behavior B: $Cx \Rightarrow (Dx \Leftrightarrow Bx)$.

But of course this is a *criterion* not a *definition*. Moreover in scientific theories, while some dispositionals (like solubility) are defined others (like electric conductivity) are not: they are taken as undefined. Likewise with psychological properties: even though actual aggression is an (ambiguous) indicator of rage, we should be ill advised to define anger as a disposition to attack, as has been suggested (Ryle, 1949) – if only because this fails to account for attacks in cold blood. Nor can we define thought as a disposition to talk, itching as a disposition to scratch, and so forth: this is just mistaking evidence for reference (Feigl, 1967). Thinking, feeling an itch and the like are just as actual as talking and scratching: that they are not public or intersubjective is another matter – one for methodology not ontology.

Dispositions are just as important as manifest properties even though the latter can explain the former, as shown by Definition 4.7. The possibility of dissolving is not the same as actual dissolution: while the latter is a property of a complex system (solute *cum* solvent), the former is a property of one of its components. For example, the solubility of salt in water consists in a certain feature of the crystal structure of dry salt which, when joined with water, gives rise to actual dissolution. We may then assert that dispositions of the causal kind are rooted in manifest properties. But this is not the same as eliminating the former. Besides, as will be seen in Sec. 5, noncausal dispositions are not so explainable: on the contrary, they help explain manifest properties. In conclusion, actualism cannot hold the fort.

3.4. *Unrealized Possibilities and Counterfactuals*

It will be noted that we have not employed counterfactuals in our elucidation of disposition. However, it is widely believed that dispositional expressions must be defined in terms of subjunctive conditionals. Thus it is commonly held that

> x is fragile $=_{df}$ If x *were* thrown against a rigid body, then x *would* break.

and

> x can do $y =_{df}$ If x *were* placed in suitable circumstances, x *would* do y.

This view seems to have originated in the operationist confusion between meaning and test: since actual breaking is evidence for fragility, and performance for ability, it is inferred that that is what they are. Science has paid no attention to this doctrine: if definable at all, i.e. if not primitive, dispositionals are defined with the help of declarative sentences only. Witness any scientific treatise or research paper.

There are good reasons for avoiding the subjunctive mood. Firstly to avoid ambiguity. Indeed one and the same subjunctive sentence may be construed in at least two different ways (Bunge, 1968d). In fact 'If A were the case then B would obtain' can be interpreted as representing either the statement

If A then B. But not A. (No inference at all.)

or the inference

If A then B. But not B. (Tacit inference: not-A.)

Such an ambiguity renders subjunctive conditionals unsuitable for scientific communication, except of course at the level of heuristic language. A second reason for avoiding subjunctive conditionals is that they are not propositions, hence they do not fall under the rules of logic. Therefore the "definitions" of fragility and ability at the beginning of this subsection are phoney.

It should not be concluded that subjunctives are junk: indeed they have considerable suggestive or heuristic power. But (*a*) they should be used with caution and (*b*) they should be avoided altogether in reconstructing scientific theories as well as in reporting experimental results. In any case, far from being capable of shedding any light, subjunctive conditionals are in need of clarification.

So much for dispositions of the causal kind. Let us now turn to a kind of disposition that is not reducible the way causal disposition is, namely chance propensity. But before doing so it will be convenient to play the probability interlude.

4. PROBABILITY

4.1. *Abstract Concept*

Let us recall the gist of the theory underlying any technical (as opposed to ordinary language) statements of the form "The probability of a equals b", or "$Pr(a)=b$" for short. Pr is a real valued function on a certain set F characterized only by its structure. To begin with F is a family of subsets of a certain abstract set S, i.e. $F \subseteq 2^S$. Further, F is closed under finite unions and intersections, and also under complementation. Therefore F is closed under finite unions and symmetric differences. So, F has a ring structure. Moreover, F is a *σ algebra* because it is a ring and $S \in F$. In other words, F is a σ algebra in the sense that its members obey the algebra of sets extended to countably finite unions. We shall call F the *probability space* or *support* of the *Pr* measure. (F is often called the *space of events* or the *sample space*, names as suggestive as they are out of place in pure mathematics.) We are now ready for

DEFINITION 4.10 Let F be a σ algebra on a nonempty set S, and *Pr* a real valued function on F. Then *Pr* is a *probability measure* on F iff

(i) *Pr* is a non-negative real valued function on F. [I.e., for every member A of the collection F of subsets of the basic set S, $Pr(A) \geqslant 0$.];

(ii) *Pr* is completely additive in F. [I.e., for any countably infinite collection of pairwise disjoint sets in F, the probability of their union equals the sum of their individual probabilities. In particular, if $A, B \in F$, and $A \cap B = \varnothing$, then $Pr(A \cup B) = Pr(A) + Pr(B)$.];

(iii) *Pr* is normed. [I.e., $Pr(S) = 1$.]

Note the following points. Firstly, the theory based on these sole assumptions is *semiabstract* insofar as it does not specify the nature of the elements of the probability space F. On the other hand the range of *Pr* is fully interpreted, hence the *semi*. This is why probability theory finds applications everywhere, from physics to metaphysics. Secondly, an obvious way in which one can satisfy the condition that F be a σ algebra, is by taking $F = 2^S$. Indeed, any power set is a σ algebra. Thirdly, the first two axioms define a *measure*, and the third makes it a probability measure. This shows that the foundations of the theory of probability constitute but a chapter of measure theory – not one of philosophy. However, such foundations, like any others, are insufficient for any applications. Indeed, as long as the probability space F is not specified, i.e. as long as no model is constructed, probability has nothing to do with possibility, propensity, or randomness. Let us dwell for a while on this important methodological point.

An *application* of any abstract or semiabstract theory to some domain of facts consists in joining to the theory two different items: (a) a model or sketch of the object or domain of facts to which the theory is to be applied, and (b) an interpretation of the basic concepts of the theory in terms of the object to which it is applied. In particular, an application of probability theory consists in joining Definition 4.10 (or some of its consequences) with (a) a stochastic model – e.g. a coin flipping model or an urn model-, and (*b*) a set of interpretation (or semantic) assumptions sketching the specific meanings to be attached to a point x in the probability space *F*, as well as to its measure $Pr(x)$. As long as these additional assumptions are not made, the theory is indistinguishable from measure theory: only those specifics turn the semiabstract theory into an application of probability theory or part of it. (For an elucidation of the notion of degree of interpretation or its dual, degree of abstraction, see Vol. 2, Ch. 7, Sec. 3.4.)

Now, before we proceed to build a model of a class of facts we must identify the latter, i.e. choose a definite ("concrete") probability space *F*. The latter may but need not represent a collection of states or of changes of state of a thing. (Whether or not a point x in *F* represents a state or an event, it is usually called either of the latter by most probability theorists.) Next we must make sure that *F* is a σ algebra, or else we must manufacture one out of the given basic set *S*. If *S* is the state space of some thing, and if we wish to speak of the probability of a state, we must make a détour: since the elements of *S* are points, not sets, we construct the power set 2^S. In other words, we construe the idiom 'the probability of state s' as "the probability of the singleton $\{s\}$". Once this algebraic matter has been attended to we can proceed to interpret the primitives *F* and *Pr*, for instance in terms of states and their propensities. Let us turn to this latter problem.

4.2. *Probability State Space*

In some cases the states of a thing can be assigned definite probabilities and some pairs of states of the same can be allotted definite conditional probabilities. As we saw in the last subsection, in order to assign probabilities to states it is expedient to construe the latter as sets. The cheapest way of doing so is to take the set of all the subsets of the set $S(X)$ of states of thing X, i.e. the power set $2^{S(X)}$, as the probability space. In this fashion we can manufacture

DEFINITION 4.11 Let X be a thing and $S(X)$ a state space for X. Then the structure $\langle 2^{S(X)}, Pr\rangle$ is a *probability state space* for thing X iff

(i) $S(X)$ is denumerable and

(ii) $Pr: 2^{S(X)} \to [0, 1]$ is a probability measure defined for every set of states and satisfying in particular the condition: $Pr(S(X)) = 1$.

Caution: the states in a probability state space are not ordered naturally according to their increasing probability, for it may well happen that different (e.g. succeeding) states have the same probability. Thus according to statistical mechanics the evolution of a closed system follows on the whole (i.e. with but a countable number of exceptions) the line of increasing vulgarity: see Figure 4.3. But the fact that there can be occasional decreases in probability values prevents us from identifying prior states with improbable states. In other words, the direction of a process is not given unambiguously by the direction of increasing probabilities.

If all the states of a thing are equally probable, the thing may be said to be *state homogeneous*. And, since no states are preferred, once the thing has attained a given state it remains there, i.e. it does not change any further. But this is of course a fiction: there are no unchanging things, hence no state homogeneous things. Real things are *state heterogeneous*, in the sense that some of their states are definitely more probable than others. In other words, a state heterogeneous thing is a thing with states that deviate from the homogeneity or equilibrium value p_0. (For a thing with a finite number N of equiprobable states, $p_0 = 1/N$.) This suggests adopting the number $h(s) = Pr(s) - p_0$ or its square as a measure of the deviance or distinction of state s. For the entire collection $S(X)$ of states of a thing X we adopt

DEFINITION 4.12 The *state heterogeneity* of a thing with state space S equals

$$H(S) = \sum_{s \in S} h^2(s), \quad \text{with} \quad h(s) = Pr(s) - p_0,$$

where p_0 corresponds to maximal (and ideal) state homogeneity.

Figure 4.4 shows three typical heterogeneity values. It will be noted that any given state heterogeneity value can be realized in diverse manners.

The following consequences follow easily because of the normalization of the total probability to unity:

(i) The state heterogeneity of a thing with N states is

$$H(S) = \sum_i p_i^2 - 1/N, \quad \text{with} \quad p_i = Pr(s_i) \quad \text{and} \quad s_i \in S.$$

(ii) As the total number of states of a thing approaches infinity, the state heterogeneity approaches the average probability:

$$\lim_{N \to \infty} H(S) = \sum_i p_i^2 = \bar{p}.$$

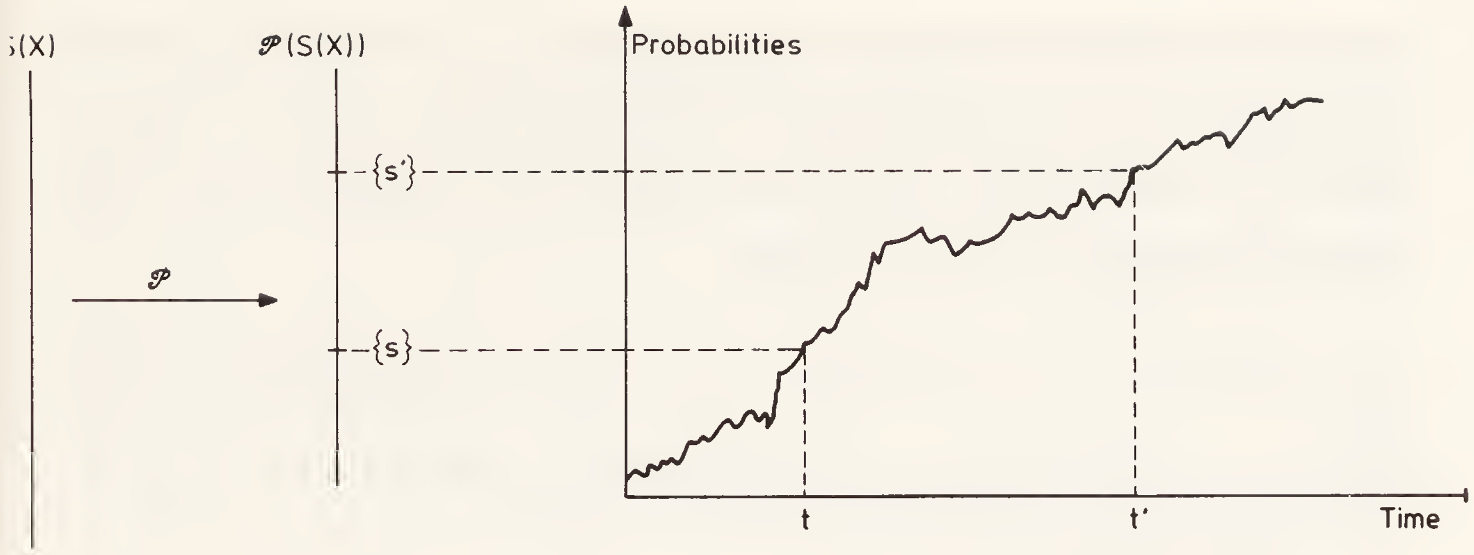


Fig. 4.3. The thermodynamic probabilities of the states of a statistical mechanical system: on the whole later stages have higher probabilities – but only on the whole, i.e. as a tendency.

The last quantity is sometimes regarded as a suitable measure of the degree of organization of a thing. This interpretation does not seem to be adequate in view of Definition 4.12. If anything, $H(S)$ measures the degree of mutability or mobility of a thing – provided S is a probability state space for the thing.

The last notion to be elucidated in this subsection is that of thing integration. Let X and X' be two different things that join to form a third thing $X \dotplus X'$. If the combination is of the conglomeration kind, the individual state spaces do not alter and the total state space is the union of the partial state spaces. And the probability that the components of the thing be in states s and s' respectively is the product of the probabilities of s and s', i.e. $Pr(s, s') = Pr(s) \cdot Pr(s')$. These probabilities change if the two components interact in the process of joining. And the stronger the interaction or correlation the greater will be the departure of $Pr(s, s')$ from the independence or noninteraction value $Pr(s) \cdot Pr(s')$.

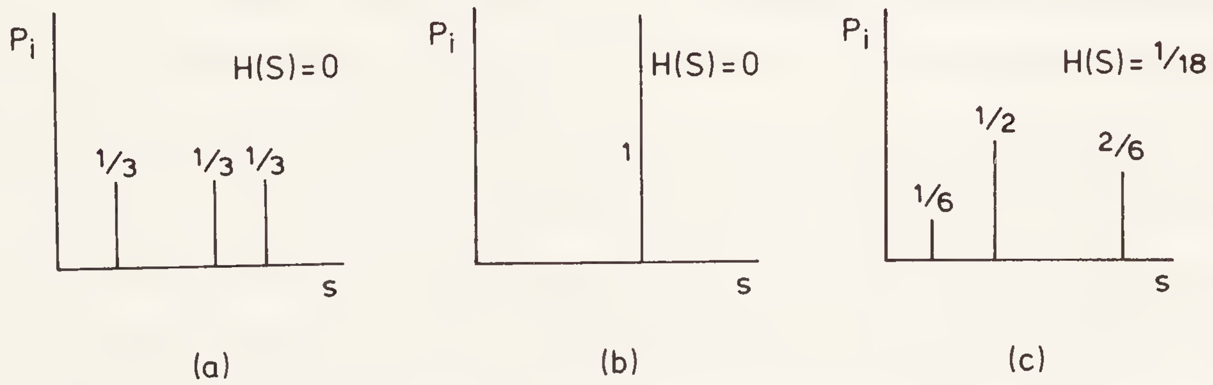


Fig. 4.4. State homogeneity (cases a and b) and heterogeneity (c).

We can then choose such a deviation, or $Pr(s, s') - Pr(s) \cdot Pr(s')$, as a measure of the interaction among the things in states s and s' respectively. The sum of the absolute values, or of the squares of these differences, for all states, seems to constitute a suitable measure of the degree of integration of the system:

DEFINITION 4.13 Let X and X' be two things with probability state spaces $\langle 2^{S(X)}, Pr\rangle$ and $\langle 2^{S(X')}, Pr\rangle$ respectively, and let $X \dotplus X'$ be the thing composed of X and X'. Then the *degree of integration* of $X \dotplus X'$ (or *strength of the interaction* between X and X') is

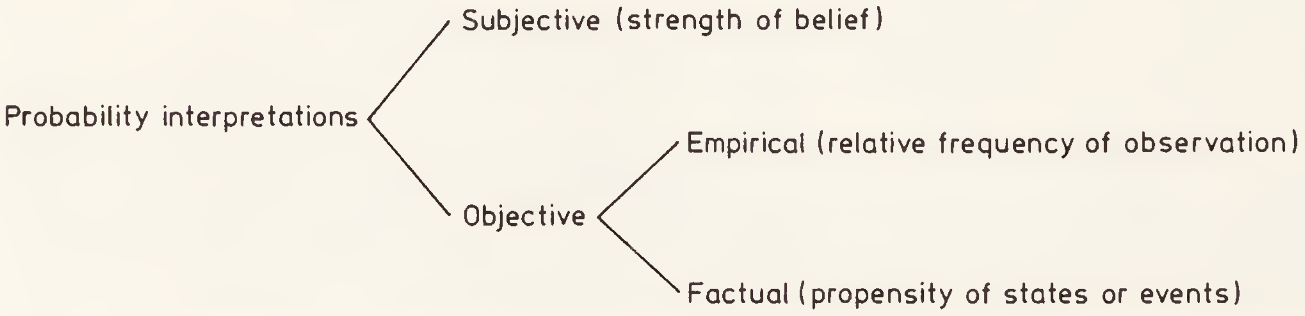
$$K(X, X') = \sum_{s \in S(X)} \sum_{s' \in S(X')} \left[\frac{Pr(s, s') - Pr(s) \cdot Pr(s')}{Pr(s) \cdot Pr(s')}\right]^2$$

provided this double sum converges.

Clearly, the above definition cannot be generalized to a system the components of which fail to have probability state spaces.

4.3. *Propensity Interpretation*

There are several interpretations of probability. The main contenders are the subjectivist (or personalist or Bayesian) interpretation and the objectivist interpretation, of which there are at least two variants: the empirical and the factual. The following tree summarizes these views.


Probability interpretations
Subjective (strength of belief)
Objective
Empirical (relative frequency of observation)
Factual (propensity of states or events)

According to the subjectivist school probability values are not objective individual properties of states of things or events in things but rather degrees of certainty about our information. Clearly, this interpretation is not the one occurring in scientific theories such as quantum mechanics or genetics, which do not concern our beliefs, let alone our beliefs about the strength of our beliefs. Science, and therefore also scientific ontology, employ an objective interpretation of probability. There are two main variants of this interpretation: the frequency and the propensity interpretations. According to the former, probability values are limits of observed frequencies: they are properties of sets of data rather than properties of the referents of such data. It has been shown repeatedly that this view is mathematically untenable, if only because frequencies fluctuate erratically around stable probability values. Strictly speaking then there is no such thing as a consistent frequency interpretation of probability: there are only frequency estimates of probability values (Bunge, 1973b, 1976a). We are then left with the strictly objective, or factual interpretation.

Recall from Sec. 4.1 that the probability calculus has two specific undefined notions: those of probability space F and probability measure Pr. And remember from Vol. 2, Ch. 6 that a full interpretation of a mathematical formalism involves interpreting all of its primitives. Now, a *factual interpretation of probability theory* consists in assigning F and every value $Pr(x)$, for $x \in F$, factual meanings. One such possible interpretation consists in regarding the basic set S, out of which F is manufactured, as the state space of a thing, hence every element of the probability space F as a bunch of states, and $Pr(x)$ as the strength of the thing's propensity to be in the state (or states) x. Similarly, if x and y are states (or sets of states) of a thing, the conditional probability of y given x, i.e. $Pr(y|x)$, is interpreted as the strength of the propensity or tendency for the thing to go from the state(s) x to the state(s) y.

Given the structure of the probability function and the interpretation of its domain F as a set of facts, the propensity interpretation is the only possible one. Indeed, if $x \in F$, then $Pr(x)$ cannot be but a property of the individual fact x. That is, contrary to the frequency view, probability is not a collective or ensemble property, i.e. a property of the entire F, but a property of every individual fact – its propensity to happen. (Hence the expression 'single case propensity interpretation', used by Giere, 1974, is redundant.) What *are* ensemble properties are derived functions such as the moments of a probability distribution (in particular its average), its standard deviation if it has one, and so on. This consideration suffices to ruin the frequency school, according to which probability is a collective or ensemble property.

As for the subjectivist (or personalist or Bayesian) interpretation, it is untenable even assuming that the domain F of Pr is a collection of mental facts. For, in the expression '$Pr(x) = y$', there is no room for a variable denoting the person who feels that the number y measures the strength of his belief in x. The personalist must interpret not only F in subjective terms: he should also have a variable representing his attitude towards his own mental facts *F*. If this variable is the generic probability value *y*, then it is the same for all persons, so there is no occasion for a subjectivist interpretation. And if it is a different variable then it does not occur in probability theory but in some psychological theory concerning uncertainty as a mental state. That different persons may come up with different estimates or guesstimates of the probability of one and the same fact is obvious but does not countenance the subjectivist interpretation any more than the uncertainties concerning electric charge values would justify a subjectivist electrostatics.

To return to the propensity interpretation. This must be distinguished from the probability elucidation or exactification of the presystematic notion of propensity, tendency, or ability. In the former case one attaches factual items to a concept, whereas in the latter one endows a factual concept with a precise mathematical structure. (Cf. Vol. 2, Ch. 6, Sec. 3.8.) In science and in ontology we need both factual interpretation and mathematical elucidation. Note also that an interpretation of probability is incomplete unless it bears on the two primitives of the theory, i.e. *F* and *Pr*. Thus the factual interpretations of Poincaré, Smoluchowski, Fréchet and others are incomplete for being restricted to assuming that a probability value is a "physical constant attached to an event *E* and to a category *C* of trials" (Fréchet, 1939). This is like saying that *e* is a physical constant without adding that *e* happens to be the electric charge of the electron. The propensity interpretation is not open to such a charge of incompleteness, for its asserts explicitly that $Pr(x)$, for $x \in F$, is the propensity or tendency for *x* to happen.

It is instructive to contrast the propensity to the actualist (or frequency) interpretations of probability values, assuming that both agree on the nature of the support set *F*. (This assumption is a pretence: not only a frequentist like von Mises but also Popper, the champion of the propensity interpretation, have emphasized that facts have no probabilities unless they occur in experimentally controlled situations.) This contrast is displayed in Table 4.3.

Note the following points. Firstly, although a probability value is meaningful – i.e. it makes sense to speak of the single fact propensity – it is so only in relation to a definite probability space *F*. Likewise a frequency value makes sense only in relation to a definite sample-population pair. For example, the formula "*x* is rare" presupposes a certain set of occurrences, to which *x* belongs, and among which *x* happens to be unfrequent.

TABLE 4.3

Potentialist vs. actualist interpretation of probability

$p = Pr(x)$	Propensity	Frequency
0	x has (almost) nil propensity	x is (almost) never the case
$0 < p \ll 1$	x has a weak propensity	x is rare
$0 \ll p < 1$	x has a fair propensity	x is fairly common
$p \doteq 1$	x has a strong propensity	x is very common
$p = 1$	x has an overpowering propensity	x is (almost) always the case

Secondly, in the case of continuous distributions, zero probability is consistent with very rare happenings: i.e. even if $Pr(x) = 0$, x may happen, though but rarely as compared with other events represented in the probability space. (All breakthroughs and all emergents – hence the most important events – have low probabilities, perhaps vanishing ones.) Consequently a fact with probability 1 can fail to happen. (Recall that the rationals have zero Lebesgue measure. Entire sets of states and events are assigned zero probability in statistical mechanics for this very reason even though the system of interest is bound to pass through them. This is what 'almost never' is taken to mean in that context, namely that the states or the events in question are attained only denumerably many times.)

Thirdly, the frequency column should be kept alongside the propensity interpretation though in a capacity other than interpretation or definition. Indeed although it fails to tell us what "$Pr(x) = y$" *means* it does tell us under what conditions such a formula is *true*. Long run frequency is in short a *truth condition* for probability statements. (For the problem whether truth conditions determine meanings, see Vol. 2, Ch. 8.)

Fourthly note again that our propensity interpretation differs from Popper's in that the latter requires the system of interest to be coupled to an experimental arrangement. No such hang-up from the frequency or empiricist interpretation remains in our own version of the propensity interpretation. Nor do we require that only events (i.e. changes of state) be assigned probabilities, as an empiricist must (since states may be unobservable): states too may be assigned probabilities, and in fact they are assigned in many a stochastic theory, e.g. statistical mechanics and quantum mechanics. (The statistical mechanical measure of entropy is a function of the probability of a state or, as Planck put it, it measures the preference [*Vorliebe*] for certain states over others.) In other words not only transition probabilities but also absolute probabilities can be factually meaningful. The requirement that only transition probabilities be considered in physics (Strauss, 1970) is untenable for the following reasons. First, since a transition probability is a conditional probability, and the latter is defined in terms of absolute probabilities – not the other way around – the former cannot have a factual meaning unless the latter has one. Second, an electron cloud (or position distribution for an electron) has a definite physical status, so much so that it can often be objectified by means of X rays.

So much for the propensity interpretation of probability. We shall see in the next section that the propensities of this kind, unlike the causal propensities, must be regarded as primary (irreducible) properties of things of certain kinds.

5. CHANCE PROPENSITY

5.1. *Irreducible Potentialities*

Classical physics assigns every point particle a definite position at each instant. On the other hand a quantum mechanical "particle" possesses at each instant a definite position distribution, i.e. a whole range of possible positions, each with a given weight or probability: see Figure 4.5. Likewise with other quantum mechanical properties, such as linear and angular momentum, spin, and energy: save in exceptional cases

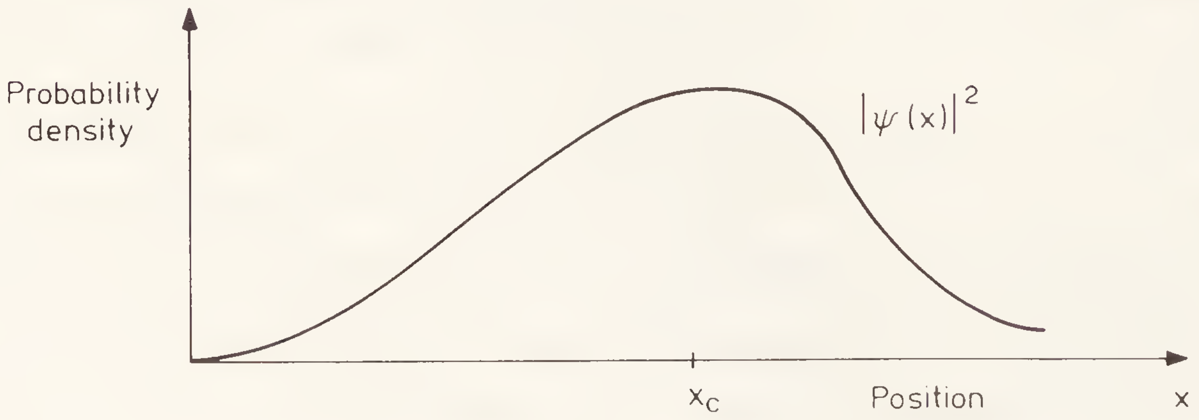


Fig. 4.5. The classical position value x_c and the quantal position distribution $\rho = |\psi(x)|^2$ of a "particle" at a given instant. The weighted average of x, i.e. $\langle x \rangle_{AV}$, coincides with x_c in the nonrelativistic limit.

each "particle" has a whole bunch (interval) of values of each of these properties. (The exceptions are constituted by the states of the "particle" that happen to be eigenstates of the corresponding property. For example, a thing in an energy eigenstate has a sharp energy value not a whole distribution of energy values. But such states, i.e. stationary states, are privileged.)

An alternative way of describing this contrast between classical and quantal properties is as follows. Where classical physics offers a nonstochastic dynamical variable Q_c, quantum physics introduces a distribution $\psi^* Q \psi$ (a "local observable"), where Q is an operator acting on the state function ψ, a place and time dependent function that determines the position probability density. (It may be assumed that these densities or bilinear forms represent the basic dynamical variables of a thing, whereas the operators Q are syncategorematic symbols serving to construct the distributions $\psi^* Q \psi$.) Every one of these random variables ("local observables") is a property of an individual thing, not a collective property of a whole bunch of similar things.

From a classical point of view this situation is intolerable. Hence the classicist will often interpret ψ as representing the state of a statistical ensemble rather than as a property of an individual thing. (This is in fact the interpretation favored by Einstein and Blokhinzev.) Yet the theory does not tolerate this interpretation of an expression such as '$\psi^* x \psi$': since it contains a single position variable x, it cannot be forced into representing a whole aggregate of particles. As long as we accept quantum mechanics we must become reconciled with its nonclassical features, such as that position is a random variable with a number of possible values, each with its own probability, for each microthing. Only the average of such a distribution is a sharp point and it coincides with the classical value. In other words, classical physics gives only a global or superficial account of microentities by ignoring the fact that the dynamical properties of the latter are random variables smeared over the entire space accessible to those entities. (See Bunge, 1977c.)

It must be stressed that the $\psi^* Q \psi$ do not just point to possible states and events: they represent properties possessed all the time by a microthing. That is, the $\psi^* Q \psi$ represent *manifest* (though not directly observable) properties rather than latent ones. On the other hand a sharp value of Q, such as any of the eigenvalues of Q, is a dispositional property of the thing concerned. In other words a thing with a Q-distribution $\psi^* Q \psi$ has the *propensity* to acquire this or that sharp Q

(eigen)value. For example, a photon travelling in free space is not concentrated at a point but, when bumping into an atomic electron, it may suddenly contract – only to be absorbed the next moment. And an electron may, under the influence of an external field, acquire a definite momentum value, though at the expense of its localization. Here the distribution is primary while the sharp values are exceptional. On the other hand in classical physics sharp values are basic and distributions derivative: they result from the interplay of numerous entities. (See Figure 4.6.) Only a classical foundation of quantum mechanics could reinstate the primacy of the sharp actual value, but then at the price of keeping the primacy of probability. (See de la Peña (1969) and de la Peña and Cetto (1975).)

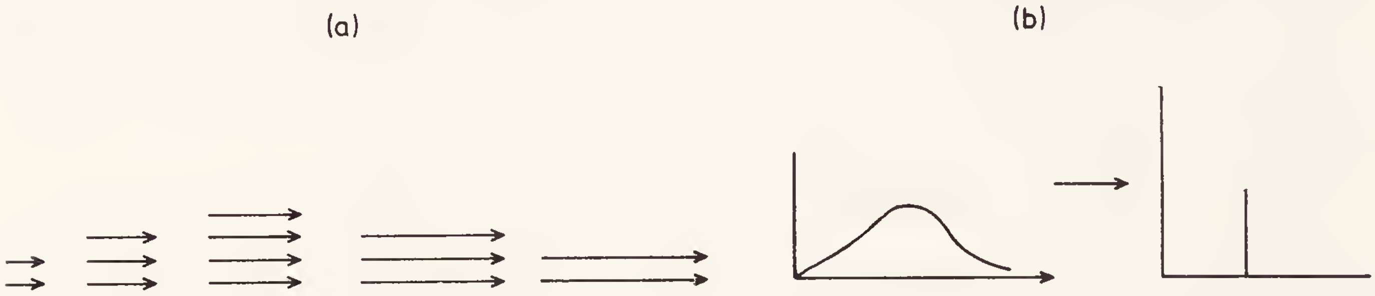


Fig. 4.6. (*a*) Classical physics: from sharp individual velocity values to distributions. (*b*) Quantum physics: from distributions to sharp individual values.

It should be stressed that the stochastic properties represented by the quantum mechanical operators Q (or their corresponding densities $\psi^*Q\psi$) are usually intrinsic properties, not properties of the thing interacting with an experimental set up. Indeed the Q's are defined as dynamical variables of the microthing of interest regardless of its circumstances: the apparatus, if present at all, has its own dynamical variables distinguished from the former by some index and acting only on the part of the state function that refers to the apparatus. The alleged ubiquity of the latter is not warranted by the mathematical formalism (Bunge, 1967b and 1973b). Besides, as emphasized by Mellor (1971), there is no point in talking of dispositions if one ascribes them to the whole "chance setup" or "experimental situation" rather than to the thing of interest, for the occurrence of a given situation is an opportunity for the disposition to display itself, not to be confused with the disposition itself. (For an excellent and detailed discussion of this matter see Settle (1974).)

5.2. *Analysis*

It will be remembered from Sec. 3 that a causal disposition or propensity is a potency that becomes actualized under suitable circumstances. The environment offers opportunities for actualization or curtails them and it is necessary and sufficient for actualization. In the case of chance propensities the possibilities lie wholly in the thing itself and the environment takes its pick. Moreover the environment may be nonexisting, in that the occurrence of an actual property may be spontaneous or uncaused, as is the case with natural radioactivity. In other words, whereas in the case of causal dispositions actualization requires that the thing of interest join another real entity, in the case of chance propensities actualization may occur if the thing joins the null thing – which is of course an artifice introduced to gain uniformity. More precisely, we make

DEFINITION 4.14 Let $z = x \dotplus y$ be a thing composed of different things x and y, where y is possibly the null thing. Further, let Q be a property of x. Then Q is said to be a *chance propensity* of x iff

(i) Q is representable nontrivially by a random variable or a probability distribution, so that each value (or interval) q of Q has a definite probability p;

(ii) y has a property R such that, when x joins y to form $z = x \dotplus y$, the probability that x acquires the individual property q (a value of Q) equals p. [I.e., $Pr(Q = q|R) = p$.]

Example 1 $x =$ electron, $y =$ diffraction system, $Q =$ position distribution of x, $R =$ letting the electron pass through a slit. *Example 2* $x =$ person, $y =$ learning environment, $Q =$ learning ability, $R =$ = allowing the person to learn.

Finally we make

DEFINITION 4.15 The *chance potentiality* of a thing x is the collection of its chance propensities:

$$\Pi_X(x) = \{P \in p(x) | P \text{ is a chance propensity}\}.$$

A thing has a greater chance potentiality than another just in case the chance potentiality of the former includes that of the latter. For example, an atom has a greater chance potentiality than any of its components; likewise a community has a greater chance potentiality than any of its members. Finally, the *potentiality* of a thing equals the union of its causal potentiality and its chance potentiality:

$$\Pi(x) = \Pi_c(x) \cup \Pi_\chi(x).$$

5.3. *Upshot*

In the present chapter we have introduced a trichotomy in the set $p(x)$ of properties of any thing x. Indeed we have assumed that $p(x)$ splits into three disjoint subsets: those of

(i) *manifest properties*, i.e. those possessed by the thing under all circumstances as long as the thing exists and remains in the same natural kind – e.g. the electric charge of an electron and the protein synthesizing activity of an organism;

(ii) *causal dispositions*, or propensities to acquire certain manifest properties under certain circumstances – e.g. the chemical reactivity of an atom and the impregnability of a mature female mammal;

(iii) *chance propensities*, or dispositions to acquire with probability certain manifest properties, depending (or not) on the circumstances.

Moreover we have drawn a radical distinction between causal and chance propensities, two categories that are usually conflated. To begin with, chance propensities are not ordinary or causal dispositions such as fragility, as there is nothing necessary about the former. In fact a distribution may collapse into any of a number of values, each possible transition having a definite probability. (Conversely, a sharp property may end up by being blurred.) Secondly, chance propensities are often basic, as shown by quantum physics and genetics. But not always: see Table 4.4. It is not true that every manifest property is rooted in a chance propensity, as extreme possibilism would have it. Nor is it true that every disposition and every chance propensity is rooted in a manifest property, as actualism claims. Both kinds of reductionism are at variance with science.

Real possibility is objective or absolute in an epistemological sense, i.e. for being independent of the knowing subject. But it is not absolute otherwise. Potencies, whether causal (like the viability of a seed) or stochastic (like the ability to learn an item on first presentation), are actualized under certain circumstances, frustrated under altered circumstances. Moreover the possibilities themselves depend upon the circumstances just as much as upon the thing concerned. Thus a peasant who migrates to a town acquires certain job and educational possibilities

TABLE 4.4

Examples of actual or manifest property, causal disposition, and chance propensity. Arrows indicate property precedence

Manifest property	Causal disposition	Chance propensity
Electric charge ———	———	→ Ionizing capacity
Sharp position value ←	———	——— Position distribution
Body mass ———	→ Body weight in a gravitational field	
Display or performance of ← an inheritable ability	——— Possession of an ← inborn ability	——— Acquisition of an inheritable ability by random gene shuffling.

that were formerly denied him, and on the other hand loses a number of traits, such as integration in a community. Hence, rather than as a monadic operator possibility should be construed as a dyadic one: fact x is possible (or impossible) under circumstance (state of the environment) y. If only for this reason modal logic is impotent in handling real possibility, whereas the relativity of possibility is incorporated into the notion of conditional probability. But modal logic deserves a separate subsection.

6. MARGINALIA

6.1. *Modal Logic and Real Possibility*

Modal logic was built in order to elucidate both the logical and the ontological concepts of possibility (see e.g. Lewis and Langford, 1932). For this reason it has often been regarded as a prerequisite for both logic and metaphysics (see e.g. Marcus, 1968). Having disposed of the former claim in Sec. 1 let us now examine the latter.

One of the most vigorous and well known defenses of the thesis that modal logic elucidates the concept of real possibility is Montague's (1960, 1974). He claimed that the principles of a certain modal calculus hold under the interpretation of '$\Box$' as "it is physically necessary that", which in turn is defined as follows: "ϕ is physically necessary iff ϕ is deducible from a certain class of physical laws specified in advance". In other words, Montague identified necessity with (theoretical) lawfulness, and disregarded circumstances. (Besides, he attributed formulas, not their referents, physical necessity, thus conflating the latter with logical necessity.) But, as we saw in Sec. 2.4, law statements, whether stochastic or not, specify only possibilities: they are so universal that they concern all possibles of a kind rather than just necessary facts. In other words, a really possible fact is one referred to by a law statement. Necessity cannot be specified without the help of the notion of circumstance, which is not incorporated into that of law. Just think of the simplest case of an equation of motion: the actual (necessary) trajectories cannot be determined unless the initial conditions are adjoined to the solutions.

Furthermore no system of modal logic is a calculus of real possibles for the simple reason that it concerns propositions not facts. For this very reason the modal logician is forced to twist "It may rain" into "It is possible that the proposition ⌜It will rain⌝ be true", or "⌜It will rain⌝ is true under some conceivable circumstances". That is, he shifts possibility from reality to statements about reality – so much more docile – thereby concealing real possibility under the rug of conceptual possibility and truth. He does the same, of course, with regard to necessity. (Worse: he has no notion of truth independent of that of satisfiability in a possible world, which he identifies with a model. And he does not know of the existence of factual and partial truth as distinct from formal and total truth.)

In science and in ontology we have little use for the notion of absolute (unconditional) possibility as well as for the modal principle ⌜$p \Rightarrow \Diamond p$⌝. What would be important is an altogether different thesis, namely: "Whatever is now the case was possible at some earlier time". But this statement includes a concept of time, which is not yet available to us and in any case is alien to modal logic. Finally ⌜$\Diamond p$⌝ is ontologically indistinguishable from ⌜$\Diamond \neg p$⌝. Thus, "That tile may fall down" has intuitively the same meaning as "That tile may not fall down" – not so in modal logic. The probabilistic versions of these statements too have the same meaning: "The probability of that tile's falling down is $\frac{1}{2}$" means the same as "The probability of that tile's not falling down is $\frac{1}{2}$" because these statements are interdeducible. Modal logic ignores all this. So we can ignore modal logic.

Most mathematical logicians ignore modal logic because they do not need it. And some, notably Quine, dislike modal logic for the wrong reason, namely because they distrust the very notion of possibility. My own evaluation of modal logic has a different rationale: I believe it is a futile game because it fails to deliver the goods it promised, namely a clarification of the notions of conceptual and of real possibility – which, to make things worse, it confuses. It is not that these notions are unimportant; on the contrary, they are much too important to be left in the care of so poor a theory as modal logic. In particular the notion of real possibility is rampant in science, which manages to handle it without the help (or rather hindrance) of modal logic. In fact every scientific theory concerns possible facts: its very referents are possible things in possible states, and every such state is representable as a point (or a region) in the state space of the thing. (To meet actuals, which theory treats as merely special cases of possibles, one has to visit observation, measurement or experiment.) But science, as we have seen, handles possibles in a way which is at variance with modal logic. For one thing in science a possible fact is a fact compatible with the laws, constraints, and circumstances (e.g. initial and boundary conditions). If the circumstances are given then the fact is either actualized or probabilified. In either case the possibility notion is construed in a fashion alien to modal logic. In other words the possibles of science are just members of some set or other – e.g. the set of nomologically possible states. Hence they are treated in a fully truth-functional way, without the assistance of modal logic. (See Figure 4.7.)

What holds for traditional modal logic holds also for the model

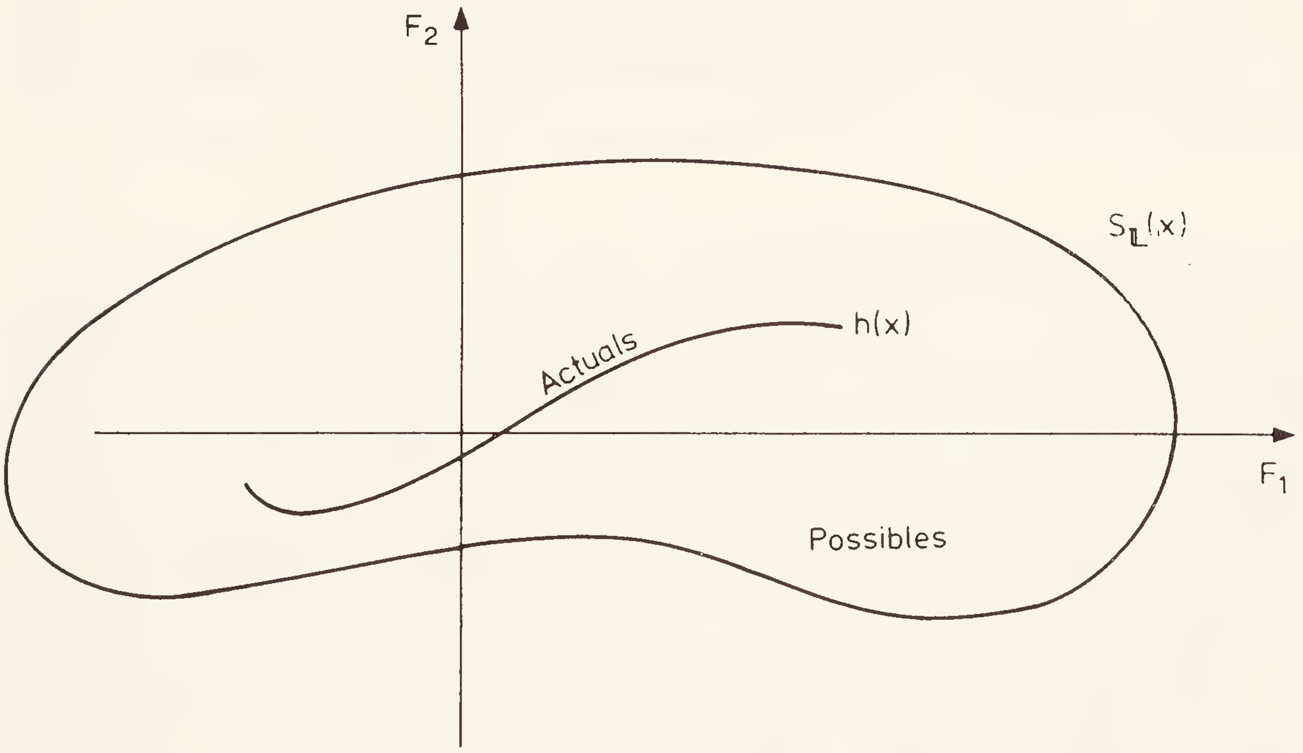


Fig. 4.7. Possibility and actuality in science and in ontology. An arbitrary point in the nomological state space $S_{\mathbb{L}}(x)$ of a thing x represents a possible state of x. The history $h(x)$ of x represents the sequence of actual states of x. No modal logic is involved.

theoretic construal of modalities (Kripke, 1959; Cresswell, 1973). According to this view possibility is satisfiability (in some model) whereas necessity is satisfiability in every model. We have admitted this idea as providing an elucidation of *one* of the four notions of *conceptual* possibility (Sec. 1.2). But the predicate "is satisfiable" refers to abstract or semiabstract (i.e. not fully interpreted) formulas – not to facts. Hence the contemporary view of modality is just as irrelevant to ontology as was Leibniz' construal of possibility as truth in some possible world.

The related notion of a model set does not fare any better with reference to ontology. In fact recall its definition (Hintikka, 1969, p. 72): "a model set [i.e. a set of formulas that are wholly true under some interpretation of the extralogical constants occurring in them] may be thought of as a partial description of a possible state of affairs or as a possible course of events ('possible world')". Here again we are concerned with abstract or semiabstract formulas not with factual law statements, let alone with factual circumstances.

In sum, neither modalities nor models for modalities elucidate the notion of real possibility. If they did, subjective idealism would be true and science redundant: we would ask the philosopher not the biologist whether sparrows could mutate into daffodils. In conclusion, modal logic fails to elucidate the concepts of real possibility – in particular those of causal disposition and chance propensity –, so it is just as useless in ontology as in semantics. Nevertheless, modal logic has promoted a new metaphysics, or rather metaphysics fiction – to which we turn presently.

6.2. *Possible Worlds Metaphysics*

If real possibility is construed as lawfulness (Sec. 2.3), then the question 'Which are the possible worlds?' becomes 'Which are the real laws?' Now, the answer to the latter is of course: 'The real laws are those inherent in the real world'. Hence the solution to the original problem is: "The set of really possible worlds is the singleton formed by the real world". Hence ours is neither the best of all possible worlds (as Leibniz and Wolff professed to believe) nor the worst of them (as the Hebrew and Christian traditions have it). Ours is the only (nomologically, really) possible world. Any alternative worlds are imaginary and belong in nonscience fiction, such as mythology. All of the (logically) possible worlds but one are (really) impossible.

Philosophers with little taste for the real world like to indulge in daydreaming about conceptually possible worlds. These provide a cozy shelter and do not have the rude manners of the real world – in particular the habit of refuting conjectures about real possibilities. A quick look at three typical problems handled by such speculations will suffice.

A first problem faced by any system of possible worlds metaphysics is of course that of defining the very concept of a possible world. This task has been performed in an elegant way by Cocchiarella (1974), who has refined, with the help of contemporary formal tools, some of the intuitions of C. I. Lewis (1923). A possible world is defined as a certain set of propositions – not of things or of facts. But this is of course a possible world of the conceptual kind (Sec. 1) – one abiding by the laws of logic not by those of factual science. Hence it is of no interest to ontology, which is concerned with the general concept of real possibility and with the question: What makes certain factual items really possible and others impossible?

As soon as one assumes the existence of a multiplicity of possible worlds one faces the question of what if anything they share. In particular, assuming that a certain individual belongs to two possible worlds, how are we to identify it? (This is the problem of transworld identification.) A special technical tool has been devised to attack this problem, or rather pseudoproblem, namely the concept of a rigid designator (Kripke, 1971). This technicality is of course unnecessary in our ontology for the simple reason that it deals only with the one real world.

Finally, some systems of possible worlds metaphysics have been applied to certain (mini)problems such as that of the counterfactuals (David Lewis, 1973). Lewis contends that "there are possible worlds other than the one we happen to inhabit" and elucidates with their help the ambiguous notion of a counterfactual. This is a typical case of what the schoolmen called an explanation of the obscure by the more obscure. Besides, who needs counterfactuals in science except as heuristic props – and even so restricted to the case when the antecedent, though contrary to actual fact, is really possible? (See Sec. 3.4.)

The following objections can be raised against possible worlds metaphysics in general:

(i) It employs modal logic, which is perfectly dispensable in all contexts except perhaps in ethics. (Whether modal logic would be needed to cope with a world different from ours we do not know. Perhaps it would be relevant in a chaotic, i.e. lawless, world – but such a world would be unlivable anyhow.)

(ii) Because it is concerned with a general concept of possibility, instead of with real possibility, possible worlds metaphysics is indifferent to worldly matters, which it is not prepared to handle. (Just think of the questions 'Is it possible to stop inflation?' and 'Is it possible for a butterfly to turn into a frog?') Hence it constitutes no ontology proper but is just an exercise in modal logic or in model theory.

(iii) Possible worlds metaphysics does not take real possibility seriously enough for, far from regarding it as inherent in the real world, it confines it to imaginary worlds. The scientific ontologist, on the other hand, has learned from quantum physics, genetics, learning theory, the theory of social mobility, and others, that possibility is inherent in reality. He has also learned that the philosopher should face the world and learn from it instead of retreating from the world by building escapist systems.

If possible worlds metaphysics is philosophically just as powerful as chess, so is the epistemology based on it. I mean the system of inductive logic where hypotheses are assigned probabilities on the basis of a count of possible worlds. Granted that counting such ghostly objects may appeal to the like of those who used to count the number of angels that could stand on the tip of a needle, one cannot help but ask two nasty questions. First: Even assuming there is a reasonable criterion for such a count, how does one go about verifying the outcome? Second: How does one allot probabilities to the various possible worlds?

6.3. *Modality and Probability*

Probability exactifies possibility, but not everything possible can be assigned a probability. There are possibilities with reference to which probability makes no sense, for the simple reason that they are not the object of a stochastic theory. For example forces may be characterized as "tendencies towards change of relative position" (Maxwell, 1892, II, pp. 211, 215). Likewise, solubility and magnetic permeability are (causal) dispositions that classical physics assigns no probabilities. Even in atomic physics, which is basically stochastic, we have similar cases. For example, a hydrogen atom at a given energy level occupies a $(2\ell+1)$-fold degenerate state, in the sense that the given energy value is consistent with $2\ell+1$ different possibilities of spatial orientation. An external field acting on the atom may remove the degeneracy by throwing the atom into any one of $2\ell+1$ nondegenerate states. This number measures the potential variety of the original (degenerate) state, or its potentiality. But as long as such possible nondegenerate states (or the possible transitions to them) are not assigned any probabilities, they remain individually nonquantified. In short, possibility does not warrant probability, be it in classical or in contemporary science.

The converse does hold of course: *wherever there is probability there is possibility*. In such cases, and in these only, "Probability is the *quantitative measure* of possibility" (Terletskii, 1971, p. 15). In other words probability provides a numerical exactification of possibility if and only if (*a*) one is dealing with chance propensity and (*b*) a stochastic theory of the facts of interest is available (Bunge, 1976a). Needless to say, a probabilistic or stochastic theory is one containing at least one random variable – i.e. a function every value of which is assigned a definite probability. In such a theory a particular probability value will occur only by postulation or by recourse to experience. For example, given a law of evolution of a probability p in the course of time, such as e.g. $\ulcorner dp/dt = kp \urcorner$, we may compute the probability value $p(t)$ at an arbitrary time t provided we give (by assumption or by observation) the initial value $p(0)$. If this additional piece of information is added, then we are allowed to make the inference to any (past or future) value of p, namely thus: $p(t) = p(0) \,.\, exp\,(kt)$. (This result is mathematically necessary but it represents only real possibility.)

In other words, we adopt the principle that anything probable according to a scientific stochastic theory T is deemed to be possible on T and moreover that the strength or weight of a possibility is equal to the corresponding probability. (Note that facts, not propositions, are assigned probabilities by scientific theories.) This idea is perfectly consistent with our nomological concept of possibility (Sec. 2.3) because, by definition, a scientific theory contains law statements (Bunge, 1967a, Chs. 6 and 7). When dealing with a stochastic theory we may use a possibility criterion more specific than the general Criterion 4.1 of Sec. 2.5 namely

CRITERION 4.2 Let T be a stochastic theory referring to a domain F of facts, and E a body of empirical data couched in the language of T and moreover relevant to T. Then if x is in F, *x is really possible according to T* and E iff $T \cup E$ contains the formula $\ulcorner Pr(x) \geqslant 0 \urcorner$.

Note that we are not equating zero probability with impossibility. Impossibility (relative to a given T-E pair) may be construed instead as the absence of a probability value. That is, anything that fails to be assigned a probability with the help of T and E will be regarded as *impossible on T and E* – though not necessarily impossible according to different premises. And anything assigned the probability value 1 with the help of T and E may be called *necessary according to T and E*. Further, two facts will be held *compossible* on T and E if the probability of their joint occurrence is defined in $T \cup E$.

We can also introduce the notion of a possibility field, namely as follows.

DEFINITION 4.16 Let F be a set of facts and let T be a stochastic theory referring to F. Then $F_0 \subseteq F$ is a *possibility field according to T* iff T defines a probability measure Pr on F_0, i.e. if, for every $x \in F_0$, $Pr(x) \geqslant 0$.

Consider now some subset F_0 of F. If the possible facts in F_0 are mutually exclusive – i.e. if the actualization of one of them precludes that of all others – they are said to constitute a *possibility fan*. In general the rays in such a fan are not equally probable. The spread of the fan, i.e. the *indeterminacy* of the bundle of possibilities, can be measured in various ways. (The best known measures are the mean standard deviation and the information-theoretic "entropy" or average information content of the bundle. However, we shall make no use of these notions here.)

In sum, whenever a set of facts constitutes a possibility field that is the concern of a stochastic theory, such a theory exactifies the coarse modal notions – not however any of the modal calculi. Table 4.5 exhibits the correspondence between the handful of modal expressions not involving iterated modalizers, and the infinitely many probabilistic statements. Out of generosity we have included the notions of near impossibility and near necessity, which are not actually handled by modal logic. On the other hand the latter makes certain distinctions, e.g. between $\Diamond p$ and $\Diamond \neg p$, and between $\Box p$ and $\Box\Box p$, which are unintelligible elsewhere.

Whereas in modal language there are only possibility and contingency between the poles of impossibility and necessity, in any probabilistic context there is a rich spectrum between them – often a nondenumerable one. Moreover what is merely possible in the short run (for a small

TABLE 4.5

Exactification of modal notions by a stochastic theory involving discrete probabilities only (i.e. excluding possibles of zero probability)

Ordinary modal language	Exact probabilistic language
x is possible	$0 < Pr(x) < 1$
x is contingent	
x is just possible	$0 < Pr(x) \ll 1$
x is almost impossible	
x is impossible	$Pr(x) = 0$
x is necessary	$Pr(x) = 1$
x is almost necessary	$0 \ll Pr(x) < 1$

sample or a short run of trials) may become necessary or nearly so in the long run. Thus the probability that a single hydrogen atom will emit a certain 21 cm line is one in 10^7 per second, yet we detect as many such waves as we wish because of the huge numbers of hydrogen atoms. The most basic physical events are extremely improbable, but they have an impact because of their sheer numbers. Give chance a chance and it will become necessity. For this reason particle physicists, atomic physicists, molecular physicists and geneticists cannot accept Borel's injunction – in the name of mathematical empiricism – to discard improbable events for being "meaningless" (Borel, 1949). Quite on the contrary, microphysicists accept the so-called "principle of compulsory strong interaction", originally formulated with tongue in cheek as follows: "Anything that is not forbidden is compulsory" (Gell-Man, 1956; see also Melvin, 1960, p. 481). A somewhat more careful wording of the above principle reads thus: All possible repetitive chance events (in particular those consistent with the conservation laws) are likely to occur in the long run. Biologists have come to accept a similar principle, namely this: Any given ecological niche where life is possible, ends up by being inhabited.

To sum up, the classical modal principles are much too poor to account for the continuum lying between impossibility and necessity. Not even the ordinary concept of probability suffices, if only because of the ambiguity of the ordinary language phrase 'x is not probable', which may stand either for "$Pr(x)$ is small" or for "Pr is not defined at x". Chance propensity can be caught only with the fine mesh of stochastic theories, which presuppose the mathematical theory of probability.

There is no better trap for catching what Hermann Weyl (1940) called "the evasive ghost of modality".

6.4. *Randomness*

Calls arrive at a telephone exchange roughly at random rather than regularly because they are made (roughly) independently from one another. However, there is nothing haphazard or chaotic about such an irregular sequence of events. In fact there is a definite probability for an arrival to occur within any given time interval. That is, the set of arrivals is what we have called (Definition 4.16) a *possibility field*. (See Figure 4.8.) Consequently a number of (probabilistic) predictions can be computed – e.g. concerning the most probable number of calls over a certain long time interval.

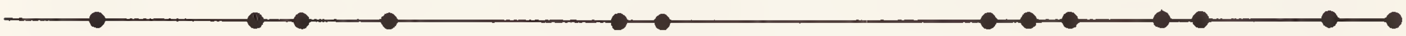


Fig. 4.8. Random arrival times – of e.g. customers at a counter, or of cosmic rays at a Geiger counter. The probability of an arrival within any given time interval is unaffected by arrivals in other intervals. See e.g. D. R. Cox, *Queues* (Methuen, London, 1961), p. 6.

In other words, because the set of arrivals has a definite probabilistic structure, a theory about it can be built; this theory is called *queuing theory*. Likewise with all other random events. For example the zig-zag walk of a drunkard is random and so is the trajectory of a pollen corpuscle immersed in a fluid, and of a particle of smoke moving in mid air. Randomness consists in this case in that there is a definite probability for each segment and each angle. Surely the precise trajectory is unpredictable. However, a number of predictions can be made for the entire set of steps because of the stochastic lawfulness of the set. For example one can predict what will be the average total distance travelled by the particle (or the drunkard) over a given time interval.

The preceding considerations suggest the following elucidation of the concept of full randomness:

DEFINITION 4.17 A set F of facts is *fully random* iff $\langle F, Pr \rangle$ is a possibility field such that the probability of any member of F is independent of the occurrence or nonoccurrence of any other possibilities in F.

If on the other hand the probability of each fact is conditional on the occurrence of some other fact, then the set of facts is not fully random.

Thus a Markov chain, though stochastic, is not fully random but is somewhere between a fully random chain and a fully determinate one. This suggests that randomness comes in degrees. In fact we can construct a measure of the degree of randomness, namely as follows. (We shall do it for a finite possibility field but the generalization to the infinite case is straightforward.) Here it is:

DEFINITION 4.18 Let F be a set of n facts and suppose that the absolute probabilities $Pr(x)$ and the conditional probabilities $Pr(y|x)$ are defined for all $x, y \in F$. Then the *degree of randomness* of F is

$$r = 1 - (1/n) \sum_{x,y \in F} |Pr(y|x) - Pr(y)|.$$

The extreme values of r are

(a) $r = 1$, i.e. *full randomness*, if the occurrence of any one fact makes no difference to that of any other fact, i.e. if $Pr(y|x) = Pr(y)$ for all $x, y \in F$;

(b) $r \cong 0$, i.e. *near determination*, if, for every x and y in F, the differences between the conditional probabilities and the absolute ones are near unity.

One might think of identifying causation with $r = 0$, since in this case whatever happens is conditioned by some other fact. ("Every event has a cause.") But this would be mistaken for the following reasons: (*a*) whereas the degree of randomness (r) is defined for facts in general, the causal relation holds only between events; (*b*) in the case of causation we do not need, and usually do not have, any probabilities; (*c*) as defined above r never becomes exactly zero.

Randomness, being a special case of stochasticity, is a type of order and should therefore not be mistaken for chaos or the total absence of law. Chaos may be defined as follows:

DEFINITION 4.19 A set of facts is *chaotic* iff no probability measure can be defined on it.

For example, the set of facts

G = {The last shot in World War II, The reader's latest dream, My first car crash, The independence of Angola} is chaotic because it is impossible to allot a probability to each member of the set in such a way that $Pr(G) = 1$ – and this because it is not possible to build a theoretical model of G.

Of course the particular set G we have just introduced to illustrate the concept of chaos or haphazardness is an artifact: the members of G do not form a natural sequence. In other words such facts cannot compound to form a real fact. We surmise therefore that chaos, though conceivable – in a nonchaotic way – is unreal. That is, none of the chaotic sequences we can imagine will be found in reality. Reality is to a large extent stochastic and even random but not chaotic, i.e. lawless.

Chaos is difficult to diagnose. Take for instance a time series or any other sequence. If the sequence is finite then, no matter how long it is, there exists at least one random variable that maps the sequence. (Think of a sequence of the form $\langle x_i \mid f(x_i) = p_i \ \& \ i \in \mathbb{N} \rangle$.) So, the sequence cannot be said to be chaotic. If on the other hand the sequence is infinite, then it cannot be drawn from observation since all experience is finite. Such a sequence must be the outcome of a calculation with some formula or algorithm. Hence it is not chaotic either. In sum even if there were chaotic sets of facts, other than those we make up for the sake of an argument, we would not be able to recognize them by their lawlessness. (See however the Kolmogoroff–Chaitin definition of a "random" – in our view chaotic – sequence in information theoretic terms: Chaitin (1974). A number sequence would be said to be completely chaotic iff the smallest algorithm capable of specifying it to a computer has about the same quantity of information as the sequence itself – i.e. if there is no general rule compressing the given information.)

6.5. *Probability and Causality*

Some philosophers, among them Suppes (1970) and Popper (1974), have offered a probabilistic elucidation of causality. It boils down to the following definition: For any events a and b,

> a is a *cause* of $b =_{df}$ The probability of b given a is greater than the absolute probability of b, i.e. $Pr(b \mid a) > Pr(b)$.

In this way those philosophers hope to reduce causal determinacy to a particular case of what may be called stochastic determinacy.

However, the reduction of causation to probabilistic correlation won't do, as shown by the innumerable counterexamples of the following kind (cf. Bunge, 1973c). The conditional probability of the reader's reading these lines, given that he was born, is greater than the absolute probability of that event, yet his birth cannot be regarded as a cause of his reading these lines. A necessary condition of course but not a cause – unless one believes in predestination. Hence causation cannot be defined in terms of probability. Both concepts are used by the scientist who computes the *probability* that a given event will *cause* a certain effect.

Whatever causes probabilifies but not conversely. Anything that probabilifies (increases chance propensities) may be called an *influence*. More precisely, if the occurrence of a state (or of a change of state) is more probable given the occurrence of another, then the latter may be said to influence the former. The concept of influence can then be elucidated in probabilistic terms provided all the state spaces of the patient are probabilistic. But this is not so in general, except of course at the level of quantum physics. Therefore we shall not adopt this definition of influence.

6.6. *The Many Worlds Interpretation of Quantum Mechanics*

We have claimed throughout this chapter that contemporary science, in particular quantum mechanics, espouses possibilism and eschews actualism. However, there is a certain interpretation of the relative state formulation of quantum mechanics (Everett, 1957) that is definitely actualist, namely the many worlds interpretation. This view, taken by DeWitt (1970), Stapp (1971), Hart (1971) and a few others, is best discussed with reference to the expansion of the state function ψ in orthogonal functions φ_n, namely thus: $\psi = \sum_n c_n\varphi_n$, where (in the simplest cases) n runs over the natural numbers. The main conflicting interpretations of this series expansion are shown in Table 4.6.

In the two main possibilist interpretations the coefficients of the expansion are probability amplitudes. In the actualist interpretation there are no probabilities: the series expansion represents the actual branching of the individual thing into infinitely many actual and noninterfering copies of it. Every branch φ_n is deemed to be actual: there is no actualization process (such as, e.g., the collapse $\psi \to \varphi_n$) for the simple reason that there is no potency to begin with. And, since these branches do not interfere with one another, there is no way of knowing that they are all simultaneously real – except of course on authority. The thing ramifies infinitely and infinitely many times, and so do any of its observers – unknown to him, as he cannot communicate with any of his copies.

TABLE 4.6
Alternative interpretations of the expansion of a state function into an infinite set of orthogonal eigenfunctions of a dynamical variable

	Possibilism		Actualism
	Operationism	Realism	(Many worlds)
ψ	Actual state before measurement	Actual state	Actual state
φ_n	Possible state before measurement	Mathematical auxiliary	Actual state
$\sum c_n\varphi_n$	Superposition of possibilities	Mathematical trick	Branching of thing into infinitely many non interfering things
$\lvert c_n\rvert^2$	Probability of finding thing in state φ_n when measuring property represented by operator	Weight of nth eigenvalue in spectrum of dynamical variable	

The obvious objections to this piece of science fiction are these. First, such infinite ramifications would violate all of the conservation laws; in particular there would not be enough energy to distribute among the infinitely many copies of any given thing. Second, the ramification is in principle undetectable, hence empirically just as irrefutable as it is unconfirmable: it takes an act of faith to accept it. Third, in the case of a dynamical variable with a nondenumerable spectrum, the world would branch into a nondenumerable set of identical copies of itself. If we wish to avoid this absurdity we must not try to interpret quantum mechanics in actualist terms and must reckon with chance propensities. We must take real possibility seriously.

7. CONCLUDING REMARKS

We have acknowledged real possibility and distinguished two kinds of it: causal disposition (e.g. fragility) and chance propensity (e.g. the propensity of an atom to occupy a definite energy level). Moreover we have regarded these two kinds of property as radically different:

whereas a causal disposition of a thing is unthinkable without a disposition in some other thing (think of the lock and key pair), a chance disposition is an irreducible property of the individual thing possessed by the latter even in the absence of an experimental set-up.

Our view on possibility is incompatible with actualism, according to which possibilities are only in the mind. We have met crypto-actualism in the guise of possible worlds metaphysics (Sec. 6.2) and open actualism in the many worlds interpretation of quantum mechanics (Sec. 6.6). A more common expression of actualism is the frequency interpretation of probability. According to the latter there is no such thing as a chance propensity for a single thing: there are only limiting frequencies defined for entire ensembles of things or for whole sets of states of a single thing, such as a sequence of throws of a coin. The phrase '$Pr(a) = b$' would then be short for something like "The relative frequency of a in a large ensemble (or a long sequence) of similar trials approaches b". This view is refuted by the existence of microphysical theories concerning a single thing, such as a single electron, to be sharply distinguished from a theory concerning an aggregate of coexisting electrons. Another example: genetics is in a position to calculate the probability of any gene combination, which, given the staggering number of possibilities, is likely to be a one time event. A relative frequency is a frequency of actuals, hence it cannot be identical with a possibility although it can measure the strength or weight of the latter. Unlike frequencies, probabilities do measure real possibilities.

If we take real possibility seriously then we must reject the narrow determinism espoused by Spinoza, Hobbes, the French *philosophes* and *géomètres*, Kant, Łukasiewicz, and many other necessitarians (as Peirce called them), to whom possibility was an epistemological category not an ontological one. (For example Kant held that all of the categories of modality express only relations to the body of knowledge: Kant (1781, p. A219). And Łukasiewicz stated that, if possibility were an ontological category, it would contradict causality and even the excluded middle principle, which he regarded as concerning actualities not possibilities: Łukasiewicz (1970, p. 35).) We have embraced instead, and tried to work out and update, the possibilism inherent in such diverse thinkers as Epicurus, Aristotle, Chrysippus, Cournot, and Peirce. However, we reject unconditional possibilism, according to which anything is possible (Naess, 1972). We have advanced rather *nomological possibilism*.

Possibilism is not only a metaphysical doctrine but, like many other ontological ideas, has implications for other disciplines, in particular action theory and ethics. Take for instance the much celebrated formula "Human freedom is the same as the knowledge of necessity", put forth by Spinoza and popularized by Engels. This formula needs correction on several counts. Firstly, by construing freedom as an epistemological category, it denies real or ontic freedom. Secondly it is obvious that knowledge, though certainly necessary to attain freedom in certain respects, is not sufficient for it. Thus the person who knows that he or she must die is not free to alter the course of events so as to avoid the final outcome. Thirdly a knowledge of real possibility is just as important as a knowledge of necessity – the more so since the latter is a particular case of the former. The person who wishes to get hold or rid of x in order to attain y will be well advised to use the bit of knowledge (or conjecture) that, in fact, the means x brings about (with a certain probability) the end y. Whether the law relating x to y is deterministic or stochastic makes little practical difference if multiple trials are allowed. What is important is that the relation exist and be known. In sum, freedom is not to be equated with the knowledge of necessity. Rather, factual knowledge of all kinds is just (and no less than) a means, among others, for attaining freedom in some respects.

So much for the concepts of possibility. We are now ready to face the general concept of change.

CHAPTER 5

CHANGE

So far we have not dealt with change. But, if we are to believe science, we must uphold the ontological postulate that all things are in flux. In fact the sciences describe, explain, predict, control or elicit changes of various kinds – such as motion, accretion, division, and evolution. Therefore ontology should analyze and systematize these various types of change.

While some metaphysicians have been busy praising change others have denied it and none have described it correctly. We shall endeavor to build a concept of change capacious enough – hence poor enough – to accommodate all of the concepts of change, particularly those occurring in the sciences, and to sketch general theories of change of some typical kinds. Such are the aims of the present chapter, devoted to change in general. The matter of qualitative change will be taken up in the companion volume, *A World of Systems*.

A change is an event or a process, whether quantitative or qualitative or both. Whatever its nature, a change is a modification in or of some thing or things: more precisely, it consists in a variation of the state of an entity. To put it negatively, there is no change separate from things – nor, indeed, are there changeless things even though some change slowly or only in certain limited respects. The world, then, consists of things that do not remain in the same state forever. This metaphysical hypothesis is an extrapolation from both ordinary experience and scientific knowledge. The contrary hypothesis, that nothing changes, is a philosophical extravagance hardly worthy of a sane person's consideration.

Since a change is a transition of a thing from one state to another, we must base the study of change on the concept of state analyzed in Ch. 3, Sec. 2. It will be recalled that a state of a thing is a list of individual properties of the thing, each of the latter being represented by a value of a state function or variable. And the collection of all possible states of a thing is called the (lawful) *state space* of the thing. Any change (of state) of a thing can be represented as a trajectory in its state space. This approach has the distinctive advantage that it does not require a detailed knowledge of the nature of the thing concerned: it is therefore ideally suited to ontology or the general theory of things. Let us then begin by recalling briefly the notions of state and state space, plunging immediately thereafter into the heart of the matter.

1. CHANGEABILITY

1.1. *Preliminaries*

Every thing is in some state or other relative to a given reference frame. And every (relative) state of a thing is uniquely determined by the properties of the latter (relative to some frame). Thus a state of a thing with n properties, each represented by a (state) function, is an n-tuple of values of the functions representing those properties. Different states correspond to (are represented by) different n-tuples. A change in the representation of properties, or in the choice of reference frame, ensues in a different representation of states. This shows that our knowledge of a state depends partly upon ourselves and the state of the art – not however that the state itself is an artifact of the beholder. States, in sum, are relative.

Pretend, for the sake of definiteness, that a certain thing x, such as a camera shutter or an electrical switch, has a single dichotomic general property: that it is either open (1) or closed (0). The possible states of x are then 0 and 1. In other words, the state space of x is $S(x)=\{0, 1\}$. Actually this is an extreme oversimplification: 0 and 1 are just the states of interest for some definite purposes. Any real thing has a large number n of properties P_i, each representable by a function F_i, where $1 \leq i \leq n$. The possible degrees or intensities of property P_i are represented as so many values of the function F_i. Hence the conceivable state space of the thing will be the cross product of the images of these functions. And the nomological state space will be a subset of the former space, resulting from the laws involving the P_i: see Figure 5.1. All of this holds, of course, for a given representation of properties, hence states.

Every point s belonging to the lawful state space $S_{\mathbb{L}}(x)$ of a thing x represents a possible state of the thing. The actual state of the thing is represented by what is called its *representative point* in the state space. An actual change of thing x is represented by a *trajectory* of the representative point; this trajectory is of course the graph of a certain function on $S_{\mathbb{L}}(x)$. It is convenient, though not indispensable, to express

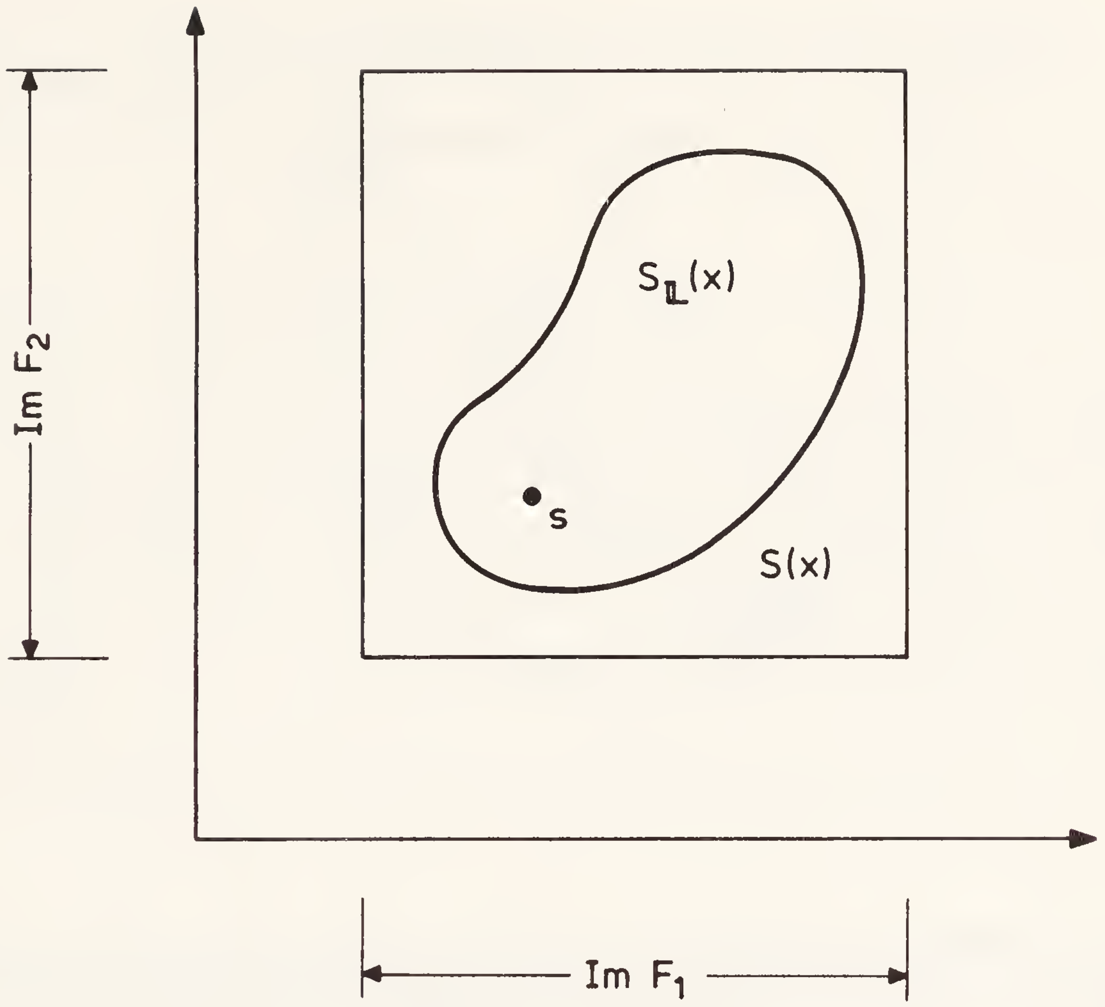


Fig. 5.1. The conceivable state space $S(x)$ and the lawful state space $S_{\mathbb{L}}(x)$ of a thing x with two properties represented by functions F_1 and F_2. '*Im F*' is read 'the image of F' and is the set of values, or range, of F.

the curve with the help of a parameter, the standard interpretation of which is time. But one of the advantages of the state space representation of change is that it requires no explicit use of the time concept.

To fix ideas consider the state space of the gene pool of a population of organisms. That set is included in the cartesian space whose axes are the gene frequencies of the pool. Every point in this space represents a possible state of the gene pool. The representative point – the one representing the actual state of the system – will remain fixed only if the organisms suffer no mutations – which is never the case, not even if the environment of the population is highly stable, such as a tropical lagoon. Otherwise the representative point moves (in state space not in ordinary space) describing a trajectory. This curve – which in the case of biological evolution has no loops – represents the genic *history* of the gene pool, or the evolution of the corresponding population at the genic level. Any segment of the total history may be called an *event* or a *process* involving the gene pool. (See Figure 5.2.)

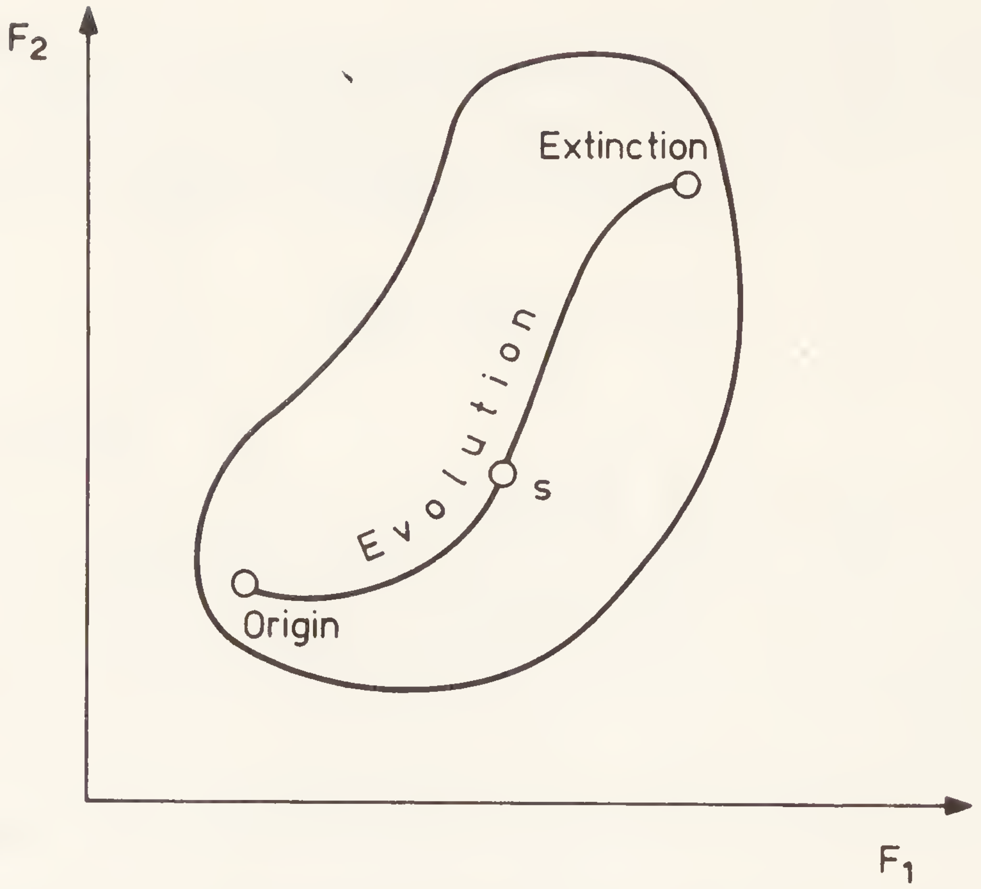


Fig. 5.2. The shifting of gene frequencies in a gene pool, in response to the joint action of mutation and selection. The end points of the curve represent the origin and the extinction of the given population.

So much for an intuitive background.

1.2. *Changeability*

Before tackling actual change we may as well elucidate the intuitive notion of mutability or changeability:

DEFINITION 5.1 Let x be a thing. Then

(i) x is *unchangeable* iff $S_{\mathbb{L}}(x)$ is a singleton for all choices of state function for x;

(ii) x is *changeable* (or *mutable*) iff $S_{\mathbb{L}}(x)$ has at least two distinct members for all choices of state function for x.

And now our first hypothesis:

POSTULATE 5.1 Every (concrete) thing has at least two distinct states, and the state space of any construct is empty. That is,

(i) if x is a thing then, for all choices of state function for x, $|S_{\mathbb{L}}(x)| \geq 2$;

(ii) if y is a construct then there are no state functions for y – or, equivalently, $S(y) = \varnothing$.

An immediate consequence of Postulate 5.1 and Definition 5.1 is

COROLLARY 5.1 All (concrete) things are changeable, and constructs are neither unchanging nor changeable.

Remark 1 We are not (yet) stating that every thing is actually undergoing some change or other, but only that it can change. *Remark 2* The second part of Corollary 5.1 states that the categories of change and of immutability do not apply to constructs. Constructs are neither eternal objects (Plato) nor changeable ones (Hegel). What do change from person to person are the brain processes occurring when constructs are thought.

Our next concept is that of overall change irrespective of the order in which it occurs:

DEFINITION 5.2 Let $S_{\mathbf{L}}(x)$ be a lawful state space for a thing x, and let $S_i(x), S_j(x) \subset S_{\mathbf{L}}(x)$ be two collections of states of x [e.g. two regions of $S_{\mathbf{L}}(x)$]. Then the *overall change* of x between $S_i(x)$ and $S_j(x)$ equals the symmetric difference between the given subsets:

$$\chi_{ij}(x) = S_i(x) \mathit{\Delta} S_j(x).$$

An overall change may be slight or large, superficial or deep. If there are considerable changes in the individual properties of a thing, but the thing neither acquires nor loses any general property, we may say that it undergoes a *large* change. If on the other hand a thing gains or loses general properties then it can be said to undergo a *deep* change. Whereas in the first case the axes of the state space remain fixed, in the second some are added or removed. We obtain a smooth description of this sort of change if we build the state space with all the necessary axes and use at any given moment only those portions of the total space that concern the properties actually possessed by the thing at the moment. In Figure 5.3 we depict the evolution of an imaginary thing undergoing a deep change. The first part of the process is described by a curve on the horizontal plane $Im\, F_1 \times Im\, F_2$; the second part is described by a curve on the vertical plane $Im\, F_2 \times Im\, F_3$. The qualitative change occurs at the junction of the two curves. The entire curve is continuous even though its tangent has a discontinuity.

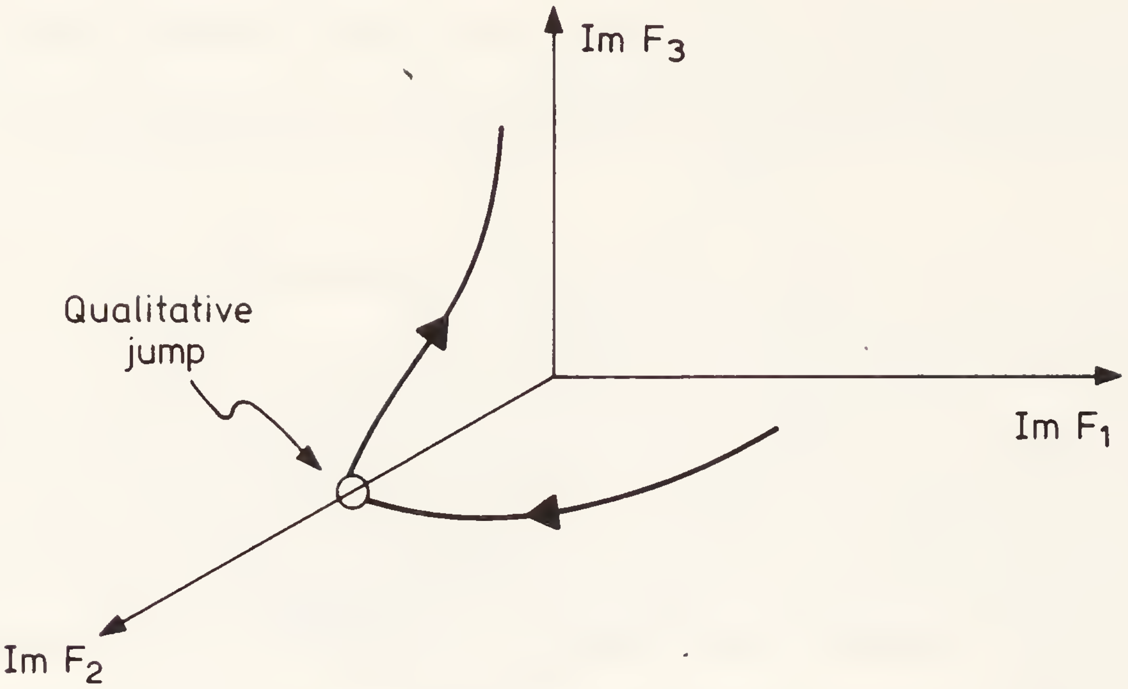


Fig. 5.3. Each arc of the curve describes a quantitative change. A qualitative jump happens at the junction of the two arcs.

In other words, we adopt

DEFINITION 5.3 Let $\mathbb{F}$ be a state function that spans the entire state space $S_{\mathbb{L}}(x)$ for a thing x. The thing undergoes a *qualitative change* iff $S_{\mathbb{L}}(x)$ equals the union of at least two subspaces, each of which is spanned by a different projection of $\mathbb{F}$. Otherwise [i.e. if none of the components can be ignored during any stretch of the process], the thing undergoes only a *quantitative change.*

An immediate consequence of this definition is

COROLLARY 5.2 Every qualitative change is accompanied by a quantitative change. [I.e., for any thing x, if x changes qualitatively then x changes quantitatively.]

The converse does not follow, which is just as well since it is false. Note that the above is not an independent ontological hypothesis but a mere corollary of a convention. The catch is that the latter has been framed so as to yield that consequence.

So much for the qualitative/quantitative distinction. The present chapter will be devoted to change in general; qualitative change will be treated in the following volume of this treatise.

2. EVENT

2.1. *The Ordered Pair Representation of Events*

Whatever changes may be thought of either as turning into a different thing or as going into a different state. In the former case the change of interest can be construed as the ordered couple $\langle x, x'\rangle$, where x and x' are the initial and the final things respectively. However, since names or singular terms such as 'x' and 'x'' are not descriptive, this representation of change is unilluminating. Besides, it forces an unnecessary multiplication of the number of things. For this reason it is not used in science and we shall not employ it in ontology.

Far more information is conveyed if the name 'x' is replaced by the sentence 'thing x is in state s', where s is a point (or a set of points) in a state space $S(x)$ for x. We can then construe a change of x as a transition from state s to some other state $s' \in S(x)$. In other words, we adopt what may be called the *principle of nominal invariance*, or permanence of names through change, and describe changes as changes of state. The principle, a designation rule, may be stated as follows:

PRINCIPLE 5.1 A thing, if named, shall keep its name throughout its history as long as the latter does not include changes in natural kind – changes which call for changes of name.

In order for this principle to work we must include in the state space of a thing all of the latter's possible states, from beginning to end. This allows us to designate a given person at age 50 by the same name as at age 5 even though in between the person may have renewed every single atom in his/her body. In short, there are no self-identical things but only constant names helping us keep track of the changes undergone by things.

As for the representation principle, a semantic assumption, it can be formulated thus:

PRINCIPLE 5.2 Let $S(x)$ be a state space for a thing x, and let $s, s' \in S(x)$ be two states of x. Then the *net change* in x from state s to state s' is representable by the ordered couple of these states, i.e. by $\langle s, s'\rangle \in S(x) \times S(x)$.

In the simplest nontrivial (yet imaginary) case, $S(x) = \{a, b\}$, and there is a single function $g: S(x) \to S(x)$ such that $g(a) = b$ and $g(b) = a$, to represent the two possible nontrivial changes of the thing, namely $\langle a, b\rangle$ and $\langle b, a\rangle$. Permanence is represented by the function $i_S: S(x) \rightarrow S(x)$ such that $i_S(a) = a$ and $i_S(b) = b$, i.e., the identity function on S.

The above principles are the clue to the theory of change to be built in the sequel. We first tackle the simpler case of the finite state space, thereafter the general case.

2.2. *The Event Space*

Let $S(x)$ be a state space for a thing x. Any pair of points in this set will represent unambiguously, though perhaps not exhaustively, a conceivable event, or change, in x. (Conceivable rather than really possible because (*a*) $S(x)$ itself is the collection of conceivable states of x, and (*b*) even if $S(x)$ is taken to be a lawful state space, not all of the transitions from or to a given state may be really possible. Just think of the state space $S(x) = \{\text{alive, dead}\}$.)

In this section we shall investigate the precise membership and structure of the space $E(x)$ of conceivable events. We start by examining the simple case of a three state thing, such as an electric switch with three states: open, closed, and transient. Calling a, b, and c the states, we have $S(x) = \{a, b, c\}$. We now form the ordered pairs

$\langle a, a\rangle$ thing x stays in state a (the identity event at a)
$\langle a, b\rangle$ thing x goes from state a to state b, etc.

Every one of these pairs represents a conceivable change in x, i.e. a conceivable event involving x. Assume further, for the sake of expediency, that in this particular case all such events can happen, i.e. are lawful. In other words, set $E_{\mathbb{L}}(x) = E(x) = S(x) \times S(x)$. We have thus 9 elementary or noncomposite events in $E(x)$:

$$\begin{array}{lll}
e_1 = \langle a, a\rangle \equiv u_a, & e_4 = \langle a, b\rangle, & e_7 = \langle b, a\rangle \\
e_2 = \langle b, b\rangle \equiv u_b, & e_5 = \langle b, c\rangle, & e_8 = \langle c, b\rangle \\
e_3 = \langle c, c\rangle \equiv u_c, & e_6 = \langle a, c\rangle, & e_9 = \langle c, a\rangle.
\end{array}$$

That is, $E(x) = S^2(x) = \{u_a, u_b, u_c, e_4, e_5, e_6, e_7, e_8, e_9\}$. The first three members of this set are nonevents or null changes. Thus u_a is the nonevent consisting in that the thing stays in state a. All others are proper events. A standard pictorial representation of this space is shown in Figure 5.4.

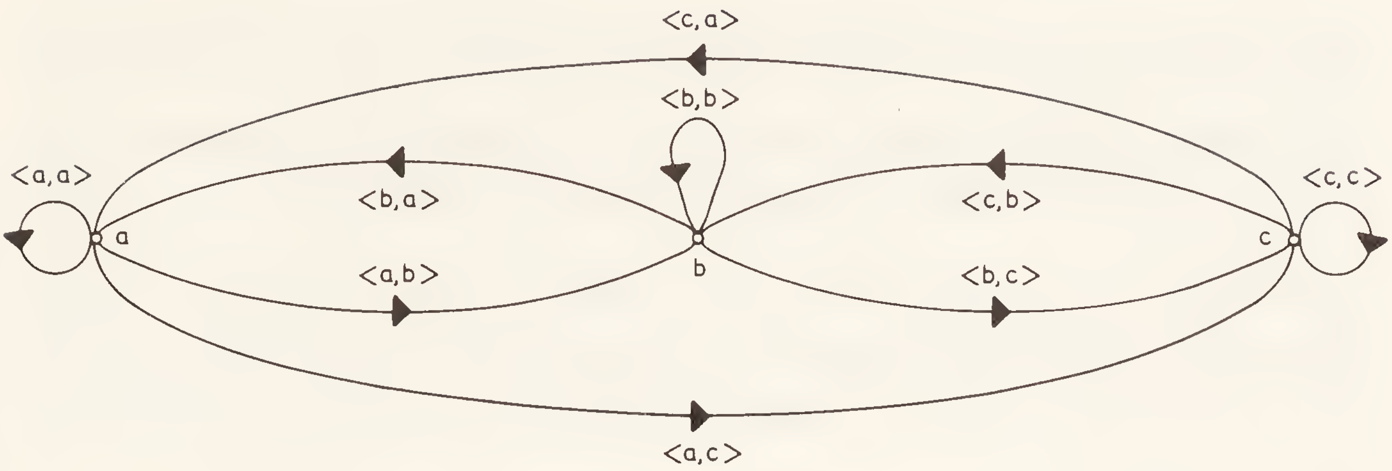


Fig. 5.4. The transition or Moore graph for a three state thing.

We assume next that the composition of certain events is possible while that of others is not. For example, event $e_4 = \langle a, b\rangle$ can be followed by event $e_5 = \langle b, c\rangle$, and the complex event may be regarded as a decomposition of event $e_6 = \langle a, c\rangle$. To symbolize the composition of events we use the asterisk $*$ and write: $e_4 * e_5 = e_6$ or, explicitly, $\langle a, b\rangle * \langle b, c\rangle = \langle a, c\rangle$. (That is, intermediate states do not show up in the net change: they are absorbed.) On the other hand the complex event $\langle b, c\rangle * \langle a, b\rangle$, which is the reverse of the former, does not occur, and so is left undefined. In other words, to represent the combination of events we introduce a (partial) binary operation $*$ in $E(x)$. If e and f are in $E(x)$, then $e * f = g$ is another event in $E(x)$ consisting in event e being *followed* by event f – with the proviso that in certain cases the composition is not defined.

The possible binary combinations of events occurring in a three state thing are shown in the following table. Every entry in this multiplication table shows the net change not the change process. But of course the

$*$	$\langle a, a\rangle$	$\langle a, b\rangle$	$\langle a, c\rangle$	$\langle b, a\rangle$	$\langle b, b\rangle$	$\langle b, c\rangle$	$\langle c, a\rangle$	$\langle c, b\rangle$	$\langle c, c\rangle$
$\langle a, a\rangle$	$\langle a, a\rangle$	$\langle a, b\rangle$	$\langle a, c\rangle$						
$\langle a, b\rangle$				$\langle a, a\rangle$	$\langle a, b\rangle$	$\langle a, c\rangle$			
$\langle a, c\rangle$							$\langle a, a\rangle$	$\langle a, b\rangle$	$\langle a, c\rangle$
$\langle b, a\rangle$	$\langle b, a\rangle$	$\langle b, b\rangle$	$\langle b, c\rangle$						
$\langle b; b\rangle$				$\langle b, a\rangle$	$\langle b,b\rangle$	$\langle b, c\rangle$			
$\langle b, c\rangle$							$\langle b, a\rangle$	$\langle b, b\rangle$	$\langle b, c\rangle$
$\langle c, a\rangle$	$\langle c, a\rangle$	$\langle c, b\rangle$	$\langle c, c\rangle$						
$\langle c, b\rangle$				$\langle c, a\rangle$	$\langle c, b\rangle$	$\langle c, c\rangle$			
$\langle c, c\rangle$							$\langle c, a\rangle$	$\langle c, b\rangle$	$\langle c, c\rangle$

table allows us to analyze certain events as processes, i.e. sequences of elementary events.

A number of regularities emerge clearly from this table. We assume that they hold for every event space and summarize them in

DEFINITION 5.4 Let $S(x) \neq \varnothing$ be a state space for a thing x and let $E(x) = S(x) \times S(x)$. The triple $\mathscr{E} = \langle S(x), E(x), *\rangle$, where $*$ is a partial (not everywhere defined) binary operation in $E(x)$ such that, for all a, b, c, d in $S(x)$,

$$\langle a, b\rangle * \langle c, d\rangle = \begin{cases} \langle a, d\rangle \text{ if } b = c \\ \text{not defined if } b \neq c, \end{cases}$$

is the *event space* of x associated with $S(x)$ iff

(i) every element of $E(x)$ represents a conceivable change of (or event in) thing x;

(ii) for any $e, f \in E(x)$, $e * f$ represents the event consisting in that event e *composes* with event f in the indicated order;

(iii) for any $s \in S(x)$, $\langle s, s\rangle \in E(x)$ represents the *identity event* (or nonevent) at s, i.e. the staying of x in state s.

An immediate consequence is that every conceivable event $\langle s, s'\rangle$, where $s, s' \in S(x)$, has a unique converse, viz., $\langle s', s\rangle$. (Which only goes to show that $E(x)$ is the totality of *conceivable* events not of really possible ones.) Another is

COROLLARY 5.3 For every state $s \in S(x)$ there is an identity event $i_s = \langle s, s\rangle$ such that, for each $e \in E(x)$ for which $*$ is defined, $e * i_s = i_s * e = e$.

In other words, the structure $\mathscr{E} = \langle S(x), E(x), *\rangle$ is a *category* with set of objects $S(x)$, set of morphisms $E(x)$, and identity morphisms i_s, for all $s \in S(x)$; the composition is $*$. The transition graph in Figure 5.6 is a vivid pictorial representation of this category in the case where $S(x)$ has only three members.

Note that in the category $\mathscr{E}$ the following holds: For any two objects (states) $s, s' \in S(x)$ there is exactly one morphism $s \mapsto s'$, namely $\langle s, s'\rangle$. This in turn implies that every morphism in the category is an isomorphism, i.e. the inverse of $\langle s, s'\rangle$ is the unique morphism $\langle s', s\rangle$. This entails

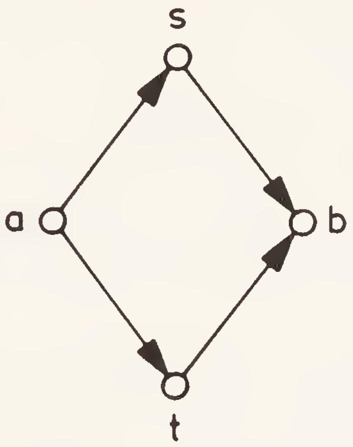
COROLLARY 5.4 Let $s, s' \in S(x)$ be states of x. Then

(i) no event other than an identity event is immediately repeatable: $\langle s, s'\rangle * \langle s, s'\rangle$ is not defined in $E(x)$;

(ii) if an event is followed by its converse then no net change results: $\langle s, s'\rangle * \langle s', s\rangle = \langle s, s\rangle$.

Hence for an event to be repeated the thing must first go back to the initial state, either through the converse event (when possible) or through intermediate events, as in the case of $\langle s, s'\rangle * \langle s', s''\rangle * \langle s'', s\rangle = \langle s, s\rangle$.

From Definition 5.4 it follows that intermediate states do not show up in the net result. Thus the two processes



are equivalent in that $\langle a, s\rangle * \langle s, b\rangle = \langle a, t\rangle * \langle t, b\rangle = \langle a, b\rangle$. Therefore we can make

DEFINITION 5.5 Two complex events (processes) in a given event space are *equivalent* iff they have the same outcome [i.e. if they relate the same initial and final states].

Hence every complex event may be construed as an equivalence class of processes, namely the set of all processes with the same end points (states).

We wind up this section by bringing to the fore two notions that have occurred tacitly in the preceding and will be of capital importance in the sequel: those of process and of precedence.

DEFINITION 5.6 A complex event [i.e. one formed by the composition of two or more events] is called a *process*.

DEFINITION 5.7 Let e and e' be two events in a given event space [i.e., e, $e' \in E(x)$ for some thing x], such that e and e' compose to form a third event $e'' = e * e'$. Then e is said to *precede* e' relative to the reference frame involved in $E(x)$:

$$\text{If } e, e' \in E(x) \text{ then } e \prec e' =_{df} e * e' \in E(x).$$

This precedence relation is irreflexive because no proper event composes with itself; it is asymmetric: if two events compose in a given order, it does not follow that they also compose in the reverse order; and it is transitive: if one event composes with another, and the latter with a third event, then the first composes with the last. In short, $<$ is a *strict partial ordering* of $E(x)$ for any thing x. (Precedence is then formally the same as $<$ in a set of numbers and $\subset$ in a set of sets.) In short, we have

COROLLARY 5.5 For any given thing x and every state representation, $\langle E(x), < \rangle$ is a strictly partially ordered set.

That every event space is ordered by $<$ means that, provided two events in it are composable, one of them precedes the other. This order is not connected: there are events in every $E(x)$ that neither precede nor succeed each other, as shown by the very definition of the composition $*$ of events as a partial operation. Also, that order is *relative* or frame-dependent instead of absolute or frame-invariant, because $E(x)$ itself is frame-dependent. Thus if e precedes e' relative to a certain frame, there may exist another frame relative to which e' precedes e – unless of course e is necessary for the occurrence of e'. Hence the importance of mentioning the space to which the events of interest belong. More on this in Sec. 2.5.

2.3. *The Representation of Processes*

The method of tracing change step by step, i.e. of collecting pairs of states, is feasible only for finite state things. (Or rather, since there are no real things with finite state spaces, we should say that the method is restricted to finite state *models* of things.) If the state space is nondenumerable like the set of reals, then there is no single state following immediately upon a given state, i.e. there is no next state. In this case, given any two states a and b of a thing, such that b follows a, there are infinitely many intermediate states following a and preceding b. In such cases then we must look for a more powerful method for representing change. We shall presently expound such a method.

In Sec. 2.1 we introduced the concept of net change. In Sec. 2.2 we noted that one and the same net change can often be effected through different intermediate states: see Figure 5.5. The net change from state s to state s' in the state space $S(x)$ can be represented by the ordered pair $\langle s, s'\rangle$. But since the change along curve f is distinct from that along

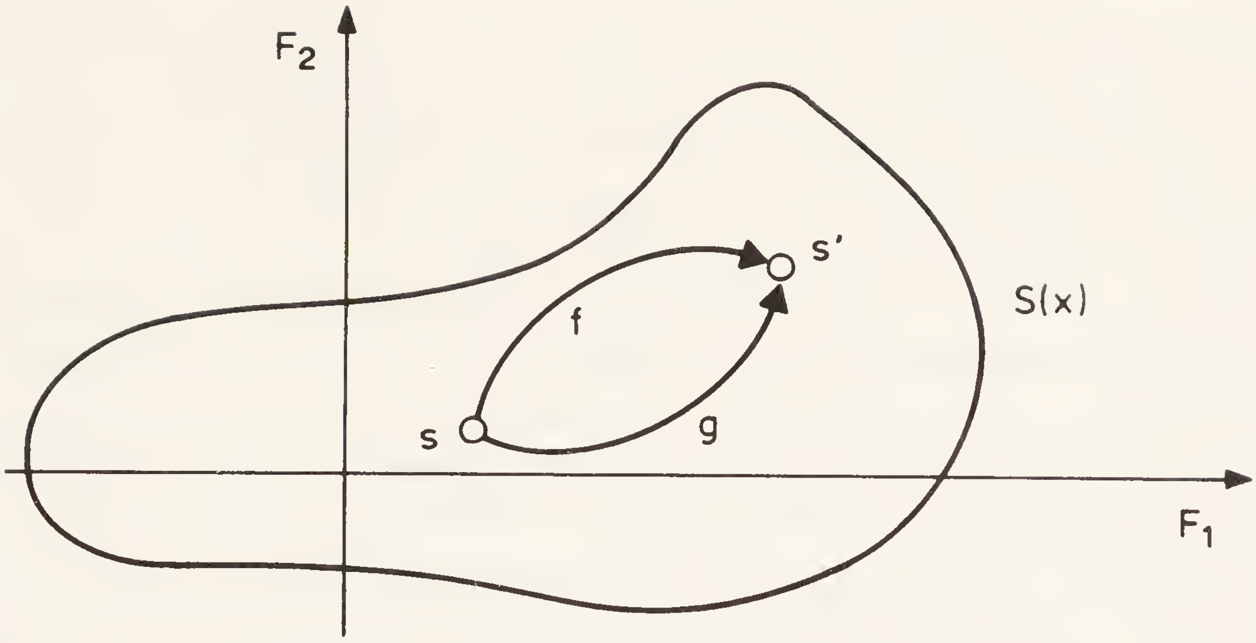


Fig. 5.5. Two different processes resulting in the same net change.

curve g, where $g \neq f$, we must represent the full events, or processes, by the ordered triples $\langle s, s', f\rangle$ and $\langle s, s', g\rangle$ respectively.

Now, the functions f and g occurring in our previous (imaginary) illustration are not supposed to be arbitrary: they must be lawful if we are to allow only lawful events and discard all the lawless (impossible) ones. In other words, f and g must be either laws possessed by the thing x concerned or transformations of the lawful state space $S_{\mathbb{L}}(x)$ compatible with those laws. (The functions need not be law functions: they may represent objective laws *cum* constraints and circumstances, such as initial conditions and boundary conditions, or they may be changes of representation leaving all the laws of the thing unchanged.) This suggests making

DEFINITION 5.8 Let $S_{\mathbb{L}}(x)$ be a lawful state space for a thing x. Then the family of *lawful transformations* of the state space into itself is the set of functions

$$G_{\mathbb{L}}(x) = \{g \text{ is a function} \mid g\colon S_{\mathbb{L}}(x) \to S_{\mathbb{L}}(x)$$
$$\& \ g \text{ is compatible with the laws of } x\}.$$

Clearly the identity transformation $i_S\colon S_{\mathbb{L}}(x) \to S_{\mathbb{L}}(x)$, where $i_S(s) = s$ for every $s \in S_{\mathbb{L}}(x)$, is a charter member of $G_{\mathbb{L}}(x)$. It represents no change and it can compose with any given nontrivial transformation.

Each such nontrivial transformation $g \in G_{\mathbb{L}}(x)$, where $g \neq i_S$, helps represent one feature of the total change undergone by x – e.g. its change of place, or of population, or of crowding. Since real changes are many-sided, they must be represented with the help of a number of different transformations composing to form the total change. Thus a thermoelectric event may be regarded as the composition of a thermal event and an electrical event – not however in the sense of 'composition' covered by the binary operation $*$ introduced in Sec. 2.2. (The acceptation of 'composition' employed in the present section is best understood by noting that $G_{\mathbb{L}}(x)$ is a proper subset of $S_{\mathbb{L}}(x)^{S_{\mathbb{L}}(x)}$, i.e. the family of all the functions on $S_{\mathbb{L}}(x)$ – not just those representing lawful changes. But the wider set has the monoid structure, since the composition of any two transformations of $S_{\mathbb{L}}(x)$ is a third transformation of the same set, and the latter includes the identity transformation.)

Now to the notion of interest:

PRINCIPLE 5.3 Let $S_{\mathbb{L}}(x)$ be a lawful state space for a thing x, and let $G_{\mathbb{L}}(x)$ be the family of lawful transformations of the state space into itself. Then a lawful event (or process) with end points s and s', where s, $s' \in S_{\mathbb{L}}(x)$, is representable as a triple $\langle s, s', g \rangle$, where $g \in G_{\mathbb{L}}(x)$ and $s' = g(s)$.

We call this the *functional representation* of change. Clearly, it subsumes the ordered pair representation, which obtains when the transformation g is fixed and the intermediate states are forgotten.

Consider the paradigmatic case of the classical model of a particle. The instantaneous dynamical state of such a (model) thing moving on a line is given by the instantaneous values of the position q and the momentum p of the particle, considered as independent variables. The nomological state space (called *phase space*) of such a thing is a subset of the cartesian plane $\mathbb{R}^2$. The transformation representing the motion can be calculated provided one knows or assumes (a) the equations of motion and (b) the forces and constraints acting on the particle. Assume, to fix ideas, that the force acting on the particle is elastic. In this case our thing is a linear oscillator and its laws are the canonical (or Hamilton) equations, which in this particular case reduce to

$$\dot{q} = p/m, \qquad \dot{p} = -kq,$$

where m is the mass of the particle and k the elastic constant. Then calling $q(t) = q$, $q(t+h) = q'$, where $h \in \mathbb{R}^+$, and similarly for the momentum, and expanding q' and p' in power series, one obtains

$$\left.\begin{aligned} q' &= q \cos ah + a^{-1} p \sin ah \\ p' &= p \cos ah - aq \sin ah \end{aligned}\right., \quad \text{with } a = (k/m)^{1/2}.$$

The value q' of the position coordinate at time $t+h$ is then a function of the earlier values of both the position and the momentum, and similarly for the momentum value p'. A transformation representing this change (motion) is called a *canonical transformation*, or a transformation satisfying the canonical equations of motion. This transformation, for a given value of the time shift h, is a map $g: \mathbb{R}^2 \to \mathbb{R}^2$ such that $g(q, p) = \langle q', p' \rangle$ as above. (There are two functions involved but they may be regarded as just two components of g.) The transformation may be regarded as a rotation in the state space, where the cartesian axes are now $x = a^{-1}p$ and $y = q$, and the rotation angle $\alpha = ah$. This transformation, far from being arbitrary, is constrained by the laws of motion. The net effect of this constraint is that the ellipse $p^2/a^2 + q^2 = 1$ is transformed into the ellipse $p'^2/a^2 + q'^2 = 1$ rather than becoming a curve of a different type.

Unless the state space happens to be one-dimensional, change will be represented by more than one function defined on the state space. Thus in the preceding example the transformation function turned out to have two components, u and v, such that $q' = u(q, p)$ and $p' = v(q, p)$. But u and v are, as noted above, the two components of a single function g mapping $S(x)$ into itself and satisfying the equations of motion.

2.4. *The Space of Lawful Events*

In Sec. 2.2 we characterized the space of conceivable net events. We now turn to a characterization of the space of really possible (i.e. lawful) events, taking into account all the intermediate states between the end points defining the net events. Recall Principle 5.3 concerning the functional representation of a full process – not just its end points – as an ordered triple $\langle s, s', g \rangle$. The collection of all such triples for a fixed transformation g may be regarded as a set of ordered couples and a subset of the space $E(x)$ of conceivable events. It represents the totality of changes of type g happening in or to thing x. And the total space of events for a given thing is the collection of all such partial event spaces. More explicitly, we propose

DEFINITION 5.9 Let $S_{\mathbb{L}}(x)$ be a lawful state space for a thing x, and let $G_{\mathbb{L}}(x) \subseteq S_{\mathbb{L}}(x)^{S_{\mathbb{L}}(x)}$ be the family of lawful transformations of $S_{\mathbb{L}}(x)$. Then

(i) the g-lawful transformations (or set of *changes of kind* g) for x, or *space of* g*-lawful events* in x, is the set of ordered pairs of states

$$E_g(x) = \{\langle s, s'\rangle \in S_{\mathbb{L}}(x) \times S_{\mathbb{L}}(x) \mid s' = g(s) \ \&\ g \in G_{\mathbb{L}}(x)\};$$

(ii) the totality of lawful transformations of x, or *space of lawful events* (in all respects) in x, is the union of all the partial event spaces:

$$E_{\mathbb{L}}(x) = \bigcup_{g \in G_{\mathbb{L}}(x)} E_g(x).$$

(We could have defined $E_{\mathbb{L}}(x)$ as the family of all $E_g(x)$'s, but then a singular event belonging to any of the latter would not belong to the former.)

Let us now relax the condition that the transformations of the lawful state space be lawful themselves, and consider the space of conceivable events. We are now in a position to generalize the result obtained in Sec. 2.2 concerning net events. Consider the events $\langle s, s', g\rangle$ and $\langle s', s'', h\rangle$, where $s, s', s'' \in S_{\mathbb{L}}(x)$, and $g, h \in G(x) = S_{\mathbb{L}}(x)^{S_{\mathbb{L}}(x)}$. That is, we consider two conceivable processes occurring in one and the same thing. Because the transformations g and h compose, the events compose in the indicated order:

$$s \overset{g}{\mapsto} s' \overset{h}{\mapsto} s''$$

to produce the net change $\langle s, s'', h \circ g\rangle$, where '$\circ$' stands for function composition. In the particular case where $h = i_S$ (identity) we get

$$s \overset{g}{\mapsto} s' \overset{i_S}{\mapsto} s'$$

which reduces to $\langle s, s', g\rangle$. This shows that the space of conceivable events is a category. More precisely, we have proved in all generality.

THEOREM 5.1 Let $S_{\mathbb{L}}(x)$ be a lawful state space for a thing x, and $G(x) = S_{\mathbb{L}}(x)^{S_{\mathbb{L}}(x)}$ the totality of (lawful and unlawful) transformations of $S_{\mathbb{L}}(x)$. Moreover let '$\circ$' designate the composition of transformations (functions) in G. Then the triple $\mathscr{E} = \langle S_{\mathbb{L}}(x), G(x), \circ\rangle$, called the *conceivable event space* of x, is a category. The objects of this category are all the states $s \in S_{\mathbb{L}}(x)$ of the thing, the morphisms all the transformations $s \overset{g}{\mapsto} s'$ with $g \in G(x)$ and $s' = g(s)$, and the identities the functions $s \overset{i_S}{\mapsto} s$.

This result (obtained with the help of Arturo Sangalli) is not terribly important because it concerns the structure of the space of conceivable events not that of the lawful events. Nevertheless it exhibits the structure in which the lawful event space $\mathscr{E}_{\mathbb{L}}(x)=\langle S_{\mathbb{L}}(x), G_{\mathbb{L}}(x), \circ\rangle$ is embedded. Besides it teaches us how to compose full events rather than just net events. In fact we have learned that the composition of events proceeds as indicated by

DEFINITION 5.10 Let $S_{\mathbb{L}}(x)$ be a state space for a thing x, and $G_{\mathbb{L}}(x)$ the corresponding set of lawful transformations of $S_{\mathbb{L}}(x)$. Further, call $e=\langle s, s', g\rangle$ and $e'=\langle s'', s''', h\rangle$ two events in the lawful event space $E_{\mathbb{L}}(x)$ defined by $S_{\mathbb{L}}(x)$ and $G_{\mathbb{L}}(x)$. Then e and e' *compose* in the indicated order to form a third event e'' iff the end of the first coincides with the beginning of the second and if the two transformations compose to form a third lawful transformation. If the composition takes place then the first event is said to *precede* the second. In symbols:

(i) $\langle s, s', g\rangle * \langle s'', s''', h\rangle = \begin{cases} \langle s, s''', h\circ g\rangle \text{ iff } s'=s'' \ \& \ h\circ g\in G_{\mathbb{L}}(x) \\ \text{undefined otherwise} \end{cases}$

(ii) $e \prec e' =_{df} e * e' \in E_{\mathbb{L}}(x)$.

And now a few overdue illustrations.

Example 1 Let the thing be a conservative system with n components. Its classical state space is the $6n$ dimensional cartesian space spanned by the state function whose components are the $3n$ generalized position coordinates and the $3n$ momenta. The event space is included in the set of graphs of the canonical or contact transformations, which are those transformations of the state (phase) space that leave the laws of motion invariant. For details on the simplest case see the example in Sec. 2.3. *Example 2* The state space of a quantum-mechanical system is a Hilbert space – a certain set of functions. The operator $T=\exp(iHt/\hbar)$ defined in this space, i.e. $T\colon S\to S$, allows one to represent the changes in the states of the system. Thus if $Q(t_0)$ represents the value of a property Q of the system at time t_0, then its value at time t is $Q(t)=TQ(t_0)T^{-1}$. *Example 3* Some or even all of the components of the state function for a thing could be random variables. Then the corresponding state space would be a probability state space (Ch. 4, Sec. 4.2.) The laws of change, or at any rate some of them, would concern not the states in a direct manner but rather their probabilities or, rather, the probabilities of the various possible (lawful) state transitions. These probabilities are conditional. Consider the simple case of a state space with a finite number of points, every one of which has a definite probability. Each final state s_f can arise from (originate in) any other state s_i with a definite probability $Pr(s_f|s_i)$, which can of course be nought for some pairs of states. For a fixed i all these probabilities add up to unity. The total probability that state s_f be actualized is

$$Pr(s_f)=\sum_i Pr(s_i)\cdot Pr(s_f|s_i).$$

2.5. *Keeping Track of Changing States*

Whereas some properties of a thing, such as its composition, are frame-invariant or absolute, others – such as its energy – are frame-dependent or relative. Since every thing possesses energy, the states of every thing are relative: recall Ch. 3, Sec. 2.8. And because states are relative, so are events. True, events such as turning on a switch or kissing look absolute. But this is only because in thinking of them we focus on what seems typical, leaving aside a number of features that are frame-dependent, such as the location and speed of the thing(s) in which the events occur. A reasonably complete account of an event calls for the determination of a number of traits, most of which are bound to be relative rather than absolute. Hence

RULE 6.1 Refer every state function – ergo every state space and every event space – to some standard thing (reference frame) or other.

The explicit reference to standards serves not only as a reminder that change is relative: it can also be utilized to label states and keep track of them in a more precise way than we have done heretofore. This is in fact the point of Definition 3.14: it allows us to identify the thing states in terms of reference states, or states of a standard. Let us recall how this was done, and let us improve on that parametrization.

Let $S(f)$ be a state space of a reference frame f for a thing x with state space $S(x)$. Further, assume that the domain of the state function $\mathbb{F}$ that spans $S(x)$, far from being unrelated to the reference frame, equals the collection $S(f)$ of reference states. That is, set $\mathbb{F}\colon S(f)\rightarrow S(x)$, so that each reference state $t\in S(f)$ is paired off to a single thing state $s=\mathbb{F}(t)=\langle F_i(t)\,|\,1\leqslant i\leqslant n\rangle$, where n is the number of properties represented by $\mathbb{F}$. The state function $\mathbb{F}$ is now the list of properties of thing x *relative to* (or parametrized by) *the states of the reference frame* f. Given any reference state t, $\mathbb{F}$ picks the corresponding thing state $\mathbb{F}(t)$. And a substitution of an inequivalent frame f' for frame f will be accompanied by different values $\mathbb{F}(t')$ of the state function.

The states of a reference frame can be described with various degrees of precision. But if we are to use them for labeling or identifying thing states, then that precision must not be less than the precision employed in describing the thing states. Now, in science one usually seeks, and often attains, maximal precision, which consists in this case in assigning real values to the individual properties of things. That is, more often than not each value of the state function F_i for a given thing equals a real number r_i, so that the state of the thing at (relative to) $t \in S(f)$ is $\mathbb{F}(t) = \langle F_i(t) \mid 1 \leqslant i \leqslant n \rangle = \langle r_1, r_2, ..., r_n \rangle$. Such a high degree of accuracy calls for a similar accuracy in the description of the reference states. And such an accuracy is achieved by coordinatizing them, i.e. by assigning them coordinate values.

The coordinatization procedure is best visualized by construing $S(f)$ as a system of four cartesian axes, so that each reference state is assigned a quadruple of real numbers. That is, we single out those properties of a reference frame that can be mapped into a four-dimensional grid, leaving all other properties aside. (The points in the grid are eventually interpreted as points in spacetime.) Finally we refer the thing states not to the reference states but to their numerical labels – as when we say '*X* was dreaming at the street corner at 7:00 p.m.'. That is, we pair each thing state off to a point in the reference grid.

The coordinatization of reference states is effected by a one to one correspondence $k : \mathbb{R}^4 \to S(f)$ between ordered quadruples of reals and reference states. The composition $\lambda = \mathbb{F} \circ k$ of the coordinatization function with the state function matches each point $r = \langle t, x, y, z \rangle$ in the reference grid with a thing state s, i.e. $r \overset{k}{\mapsto} t \overset{\mathbb{F}}{\mapsto} s$, as shown in Figure 5.6 and in the commutative diagram overleaf.

Now, the reference grid is partially ordered. Indeed, if r and s are quadruples of reals, then $r \leqslant s$ iff each component or coordinate of r is smaller than the corresponding coordinate of s. That is,

$$\langle r_1, r_2, r_3, r_4 \rangle \leqslant \langle s_1, s_2, s_3, s_4 \rangle \text{ iff } r_i \leqslant s_i \text{ for all } i = 1, 2, 3, 4.$$

Consequently $\langle S(f), \leqslant \rangle$ is a partially ordered set. This order is not connected: not every pair of reference states is related by $\leqslant$: see Figure 5.7. And the same kind of order is imposed, via $\mathbb{F}$, on the set $S(x)$ of thing

$$\begin{tikzcd}
\mathbb{R}^4 \arrow[r, "k"] \arrow[dr, "\lambda"'] & S(f) \arrow[d, "\mathbb{F}"] \\
 & S(x)
\end{tikzcd} \qquad \lambda = \mathbb{F} \circ k$$

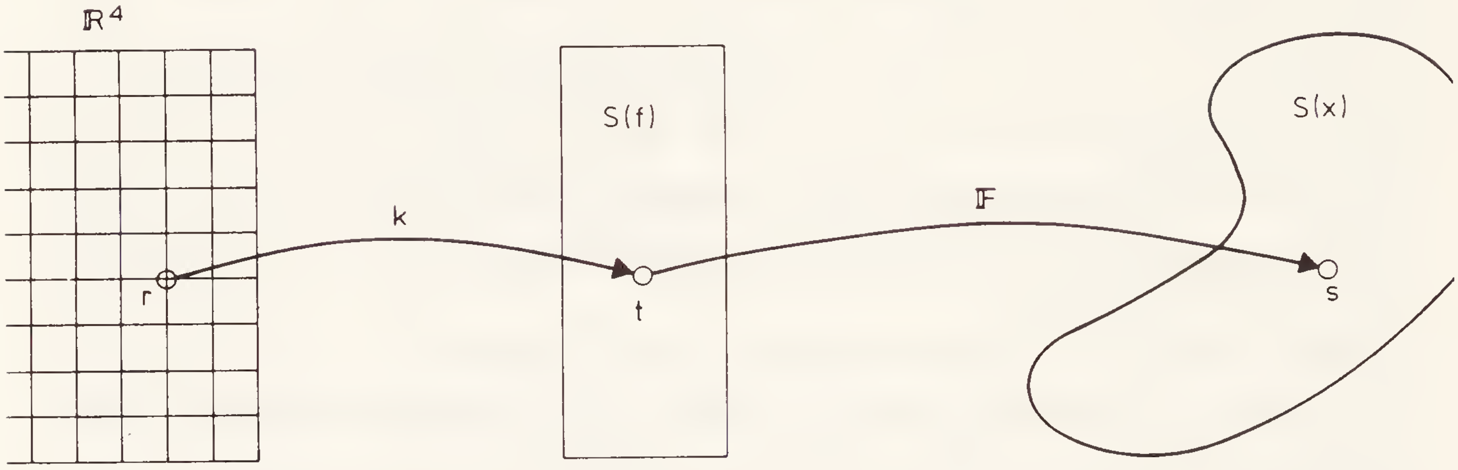


Fig. 5.6. Coordinatization of states. Each reference state $t \in S(f)$ is assigned at most one thing state $s \in S(x) \subseteq \mathbb{R}^n$ and is represented by a quadruple of reals $r \in \mathbb{R}^4$.

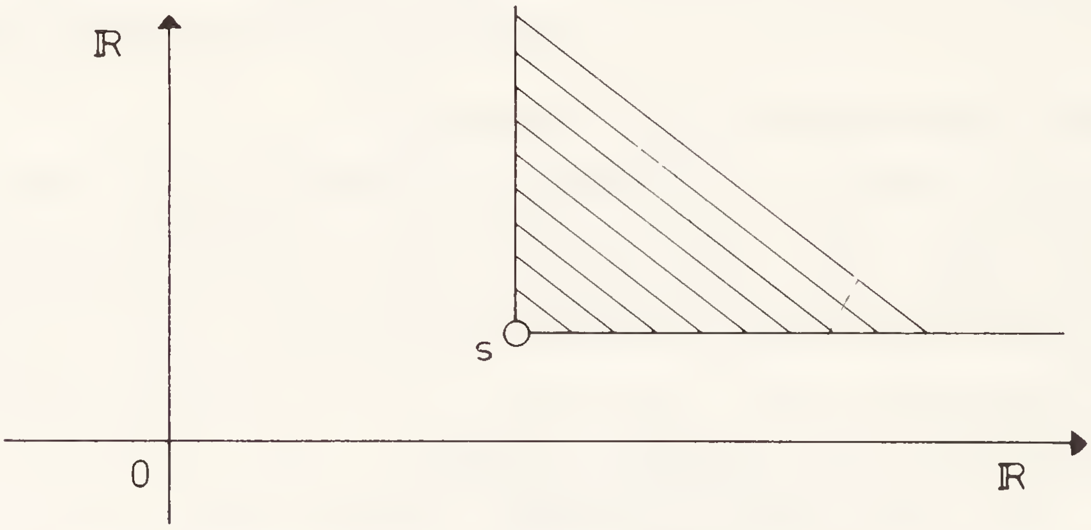


Fig. 5.7. State s precedes all others in the shaded region but none outside the latter.

states even if $\mathbb{F}$ does not take values in $\mathbb{R}^n$ – as when a mammal is said to pass through four stages of development.

The preceding considerations can be compressed into

DEFINITION 5.11 Let f be a reference frame for a thing x and let $S(f)$ and $S(x)$ be the respective state spaces. Further, assume that $S(f)$ is spanned by a four-component state function, while $S(x)$ is scanned by an n-component state function $\mathbb{F}$. Then a *coordinatization* of the states of both x and f is a $1-1$ correspondence $k: \mathbb{R}^4 \to S(f)$ such that

(i) k assigns each reference state $s \in S(f)$ a quadruple of real numbers r – as a consequence of which each thing state $s \in S(x)$ is identified as the n-tuple $s = (\mathbb{F} \circ k)(r)$;

(ii) k is order preserving, i.e. the partial order of $\mathbb{R}^4$ induces via k the partial ordering of reference states, which in turn orders the states of the thing: for all $t_1, t_2 \in S(f)$,

$$\begin{aligned}&\text{if } r_1 \leqslant r_2 \quad \text{then} \quad k(r_1) \leqslant k(r_2) \ \& \ \mathbb{F}(t_1) \leqslant \mathbb{F}(t_2),\\ &\text{where} \quad t_i = k(r_i).\end{aligned}$$

In other words, a state coordinatization is an order preserving function labeling and ordering the thing states via the reference states:

$$\langle \mathbb{R}^4, \leqslant \rangle \xrightarrow{k} \langle S(f), \leqslant \rangle \xrightarrow{\mathbb{F}} \langle S(x), \leqslant \rangle.$$

As the reference frame passes from one state to another, so does the thing: the successive states of a thing can then be matched to, hence identified by (not *with*), the successive states of the standard. And in turn, the ordering of states induces an ordering of events (in the same thing and relative to the same frame). Hence we can make

DEFINITION 5.12 Let s, s', s'', s''' be four states of a given thing relative to a certain reference frame, such that the first two form a net event $e = \langle s, s' \rangle$ and the last two another event $e' = \langle s'', s''' \rangle$. Then

(i) $e \preccurlyeq e' =_{df} s \leqslant s'' \ \& \ s' \leqslant s'''$;

(ii) $e \sim e' =_{df} e \preccurlyeq e' \ \& \ e' \preccurlyeq e$.

The relation $\preccurlyeq$ of event order is reflexive: two events are congruent iff their corresponding initial and final states are congruent, i.e.

$$e \sim e' =_{df} s \sim s'' \ \& \ s' \sim s'''.$$

It is also clear that $\preccurlyeq$ is antisymmetric and transitive. Hence $\preccurlyeq$ is a partial order – unlike the relation introduced by Definition 5.7, which was only a strict partial order. Hence instead of Corollary 5.5 we have

COROLLARY 5.6 For every thing x and relative to any given reference frame, $\langle E(x), \preccurlyeq \rangle$ is a partially ordered set.

Our new definition of event order is more comprehensive than Definition 5.7, as it includes the particular case when e composes with e' to form a third event $e'' = e * e'$ (in which case $s' = s''$), as well as the case when they do not compose. (See Figure 5.8.) The composition of events is possible only if the precedence is proper: $e * e' = e''$ iff $s' = s''$ & $e < e'$.

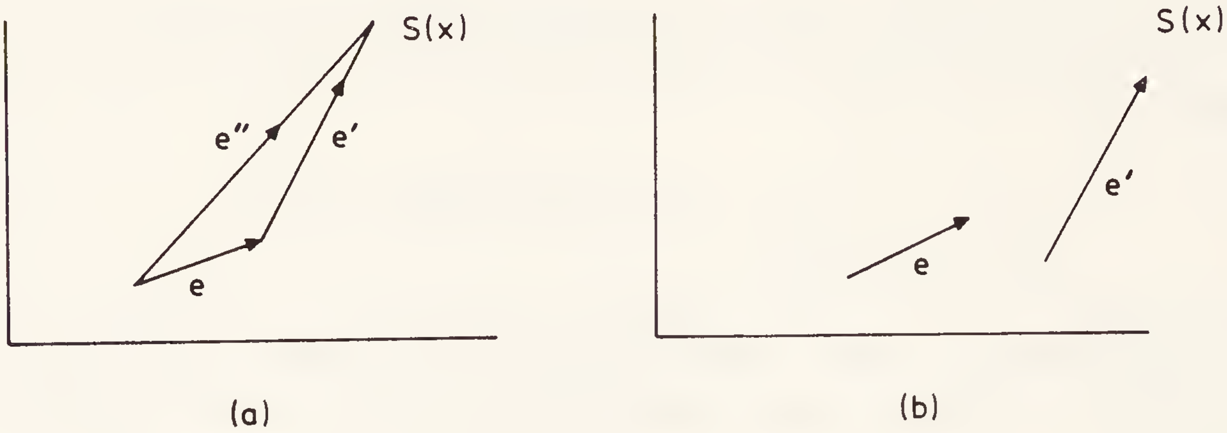


Fig. 5.8. Event order. (*a*) $e < e'$ and both compose to form e''. (*b*) $e < e'$ but they do not compose.

Consequently two events do not compose if neither precedes the other. Which is rather obvious, as the links in a chain of events are supposed to follow one another. Which in turn invites defining a *chain of events* as a denumerable and strictly partially ordered set of events $E^*(x)$ included in the event space of a thing x such that, for any $e, e' \in E^*(x)$, if $e < e'$ then the end state of e coincides with the beginning state of e', so that both events compose. More on this in Sec. 3.1.

Finally we shall proceed to formalize the notion of a variable collection, i.e. one whose membership changes as the reference states succeed one another – e.g. the population of a town. Obviously the concept of a set won't do because every set has a fixed membership. But instead of using a single set we may employ a family of sets indexed by a change parameter – call it t. Thus calling C_t a collection at t, we form the family $C = \{C_t | t \in T\}$, which may be called a variable collection. Or, equivalently, we may propose

DEFINITION 5.13 Let $t \in S(f)$ be the change parameter of frame f. Then any function

$$K: S(f) \rightarrow C, \text{ with } C = \{C_t | t \in S(f)\},$$

such that $K(t) = C_t$ for $t \in S(f)$, is called a *variable collection.*

This construal of a variable collection will be adopted henceforth, particularly in Vol. 4 of this treatise when dealing with populations of molecules, organisms, or persons – the membership of which, unlike that of a set, changes in time. It is particularly advantageous for analyzing properties of collections. Take for example the numerosity or population of a variable collection such as an ecosystem. It may be regarded as a function $p: C \to \mathbb{N}$ assigning each member of the family C a natural number $n \in \mathbb{N}$. This function composes with K to yield the function $f = p \circ K: S(f) \to \mathbb{N}$, which assigns to each reference state (or instant) $t \in S(f)$ a natural number, namely the population of the variable collection. (For discrete times the mathematical theory of sets through time is available.)

2.6. *Rate, Extent, and Change Potential*

A first notion we want to clarify is that of rate of change. Actually there is a large family of concepts of rate of change, of which we shall present only the two most common ones.

Suppose both the reference states and the thing states have been duly coordinatized (Sec. 2.5). Every reference state $t \in S(f)$ is then represented by a quadruple of reals: $t = \langle \alpha, \beta, \gamma, \delta \rangle$. Consequently the corresponding thing state is $s = \mathbb{F}(t) = \mathbb{F}(\alpha, \beta, \gamma, \delta)$. If $\mathbb{F}$ has a component F_i that is once differentiable with respect to one of the four reference parameters, say α, then the rate of change of F_i with respect to α is the partial derivative of F_i with respect to α. The thing does not change at all in the ith respect over the interval $[\alpha_1, \alpha_2]$ iff $\partial F_i / \partial \alpha = 0$ for all $\alpha \in [\alpha_1, \alpha_2]$. It changes swiftly in the same respect over the same interval if the corresponding rate of change is appreciable relative to F_i itself – otherwise it changes slowly. Needless to say, the reference parameter may but need not be the time. In sum, we can make

DEFINITION 5.14 Let the states of a thing x be coordinatized through those of a certain reference frame, so that each thing state is of the form $s = \mathbb{F}(\alpha, \beta, \gamma, \delta)$, where the arguments are the parameters identifying the reference states. Further, let the ith component of $\mathbb{F}$ be differentiable with respect to α. Then

(i) the *relative rate of change* of x in the ith respect and with respect to α is

$$V_i = \frac{1}{F_i} \frac{\partial F_i}{\partial \alpha};$$

(ii) the *relative extent of the change* of x in the ith respect, with respect to α and over the interval $[\alpha_1, \alpha_2]$, is

$$\Delta_i(\alpha_1, \alpha_2) = \frac{\ln F_i(\alpha_2) - \ln F_i(\alpha_1)}{\alpha_2 - \alpha_1} \quad \text{provided } F_i(\alpha_1), F_i(\alpha_2) > 0.$$

If the thing changes periodically with period $[\alpha_1, \alpha_2]$, the extent of its change over an integral multiple of the period is nil. This provides a definition of the concept of *cyclic change*.

The preceding definition is of no use in the case of a probability state space the points of which are not paired off to any reference frame. In this case we are interested in global changes that are frame invariant. The rate of change from state s_i to state s_j may be taken to be the conditional probability of s_j given that s_i has been actualized, i.e. $Pr(s_j|s_i)$. The rationale behind the choice is this. Think of a complex thing, n_i components of which are in state s_i, n_j in state s_j, and so on. Then $n_i Pr(s_j|s_i)$ will measure the most probable number of components jumping from state s_i to state s_j. The larger this product the greater the rate of change. The stochastic matrix $\|Q_{ij}\| = \|Pr(s_j|s_i)\|$ represents then the global rate of change of the thing. If the matrix is diagonal, i.e. if $Pr(s_j|s_i) = \delta_{ij}p_i$, where δ_{ij} is the Kronecker delta and p_i the probability of state s_i, then the thing does not change. The larger the change the larger the corresponding off-diagonal terms of $\|Q_{ij}\|$. And the extent of change, also called the *mobility*, equals the sum of the products of the off-diagonal terms by their symmetrical counterparts. We compress all this into

DEFINITION 5.15 Let $\langle S(x), Pr\rangle$, where $S(x)$ is a denumerable state space for a thing x, and Pr a probability measure on $S(x)$, be a possibility field for x. Then

(i) the *global rate of change* of x is the stochastic matrix

$$\|Q_{ij}\| = \|Pr(s_j|s_i)\| \quad \text{with} \quad s_i,\ s_j \in S(x),$$

where $Pr(s_j|s_i)$ is the probability of state s_j given that state s_i has been actualized;

(ii) the *extent of change* (or *mobility*) of x is

$$\Delta = \sum_{i \neq j} Q_{ij} \cdot Q_{ji} = \sum_{i \neq j} Pr(s_j|s_i) \cdot Pr(s_i|s_j).$$

The latter is a measure of the changeability of certain things in certain respects, namely those describable with the help of denumerable possibility fields. But it is not a measure of the overall changeability of an arbitrary thing. Just how changeable (in all respects) is an arbitrary thing? Equivalently: How big are event spaces? To answer this (imprecise) question consider the set of *all* properties of any real thing and a realistic theoretical model of such a thing – say, the relativistic theory of the electron, or of the gravitational field. Such a model, or functional schema, is based on a certain set on which an n-component state function $\mathbb{F}$ is defined, every component of which represents a (changeable) property of the thing. At least some of the components of the state function are bound to be continuous with respect to some of the "independent variables". Call F_i any such continuous state variable. As the thing changes, the values of F_i vary continuously (or at least piece-wise continuously) even if those of some other components of $\mathbb{F}$ change by jumps. (Think of a thing that gains or loses quanta of electric charge, or people, while losing or gaining momentum.) See Figure 5.9. Clearly, the (state) space spanned by $\mathbb{F}$ is nondenumerable.

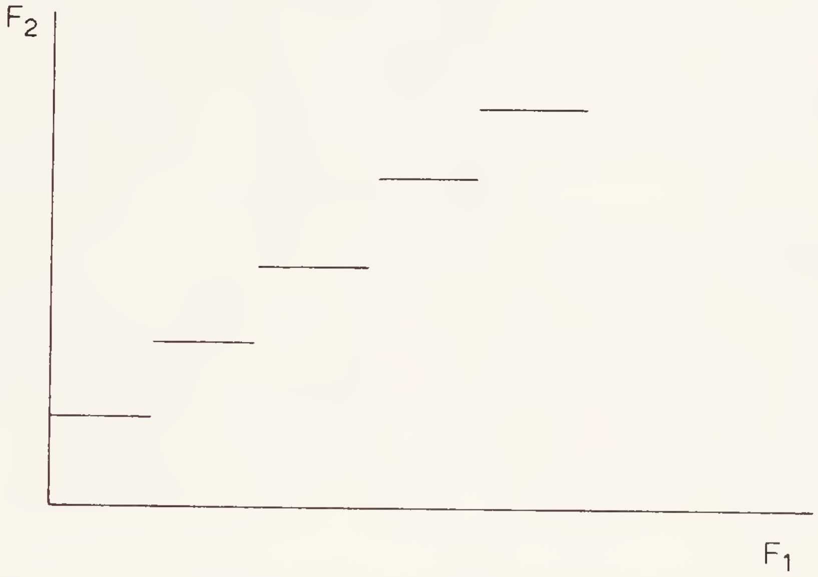


Fig. 5.9. A state space with a continuous axis (F_1) and a discontinuous one (F_2). The broken line represents the evolution of the thing. The totality of states is nondenumerable.

We assume that every thing has at least one property that changes continuously with respect to some variable or other – the "independent variable" being usually a time coordinate or a space coordinate:

POSTULATE 5.2 Every [realistic] state function has at least one continuous component.

Since a state space is the set of points scanned by the tip of the state function, we have

COROLLARY 5.7 Every [realistic] state space is nondenumerable.

And from the way events are built we also infer

COROLLARY 5.8 Every [realistic] event space is nondenumerable.

True, we often build models of things (model objects) whose state spaces are denumerable or even finite. But this is done for the sake of expediency rather than truth – i.e. because we are interested in just a few aspects of the whole thing. Thus, even though automata theory models certain machines as having only a finite number of states, the actual design and construction of such machines presupposes theories involving nondenumerable state spaces such as those analyzed in Ch. 3, Secs. 2.4, 2.5. Even if the universe were to consist of just two things in relative motion (e.g. one diatomic molecule), the world event space would be nondenumerable. In conclusion, realistic event spaces are as big as any subset of the real line can be: they all have the power of the continuum.

Because all realistic event spaces are equally bulky (Corollary 5.8), they do not supply a suitable measure of the change potential of real things. We must look elsewhere for such a measure. If we scan the factual sciences in search for a property shared by all things, regardless of kind and related to their changeability, we find but one universal property – namely, energy. Moreover energy has the property that it is greater the more complex the thing, hence the greater its potential for change. In fact, energy is strictly additive for noninteracting things, i.e. when the total state space of the compound equals the union of the partial state spaces (recall Ch. 3, Sec. 2.7). We therefore make

POSTULATE 5.3 There is exactly one property possessed by all things, that is additive for thing aggregates composed of detached parts. That is, if Θ is the class of things and F that of reference frames, then there exists one and only one real valued function $\varepsilon : \Theta \times F \to \mathbb{R}$ such that, for any x and y in Θ such that $x \rceil\!\!\lfloor y$,

$$\varepsilon(x \dotplus y, f) = \varepsilon(x, f) + \varepsilon(y, f) \text{ for all } f \in F.$$

This function ε is called *energy*, and its value at $\langle x, f \rangle$ – namely *the energy of x relative to f* – measures the *changeability of x relative to f*.

(*Energeia* was an ontological concept before being absorbed and transformed by modern physics. It is still sometimes regarded as a metaphysical notion though for the wrong reason, namely because it does not represent a perceptible property. The general concept of energy *is* a genuine ontological notion for a different reason, namely because it designates a universal property. For this reason it occurs everywhere in science, even though in some areas – particularly geography and history – it has entered only recently. In any event it is nice to see the concept of energy back in the philosophical fold. Because we now understand that energy is a property of things, there is no danger of it being reified, let alone be made into the only existent, as it was by the energetists of the turn of the century – e.g. Ostwald, 1902.)

It may be objected that Postulate 5.3 is not precise enough to define the energy concept. To this it may be rejoined, first, that science does not supply a definition of the general concept of energy. Second, to a physicist Postulate 5.3 is enough to *identify* ε as the energy function, for it is the only quantitative (and approximately additive) property all things have. (For things with interacting parts, the energy is slightly subadditive.) All our axiom states is that *all things have energy*. It does not supply a method for computing the energy of any thing in particular: this is a task for science. And no science can accomplish this task without some precise information (or conjecture) about the kind of thing, its structure, interactions, and circumstances.

One can conceive of certain properties that come in infinite degrees. For example, if the universe happens to be infinite in extension, then there are things that are infinitely distant from one another. On the other hand the energy of a finite part of the universe is normally taken to be finite. (True, both in classical and in quantum physics there are theories containing theorems – never axioms – according to which certain things, such as electrons and plane electromagnetic waves, have infinite energies. However, these formulas are regarded as blemishes of the theories not of nature, and several tricks are resorted to in order to avoid them.) We are therefore justified in adopting

POSTULATE 5.4 The energy of any thing composed of a finite number of things, and relative to any reference frame, is finite.

An important consequence of the finite energy (or potential for change) of every finite thing is that the pace and the extent of every change are bounded. (See Figure 5.10.) For example, the speed of a

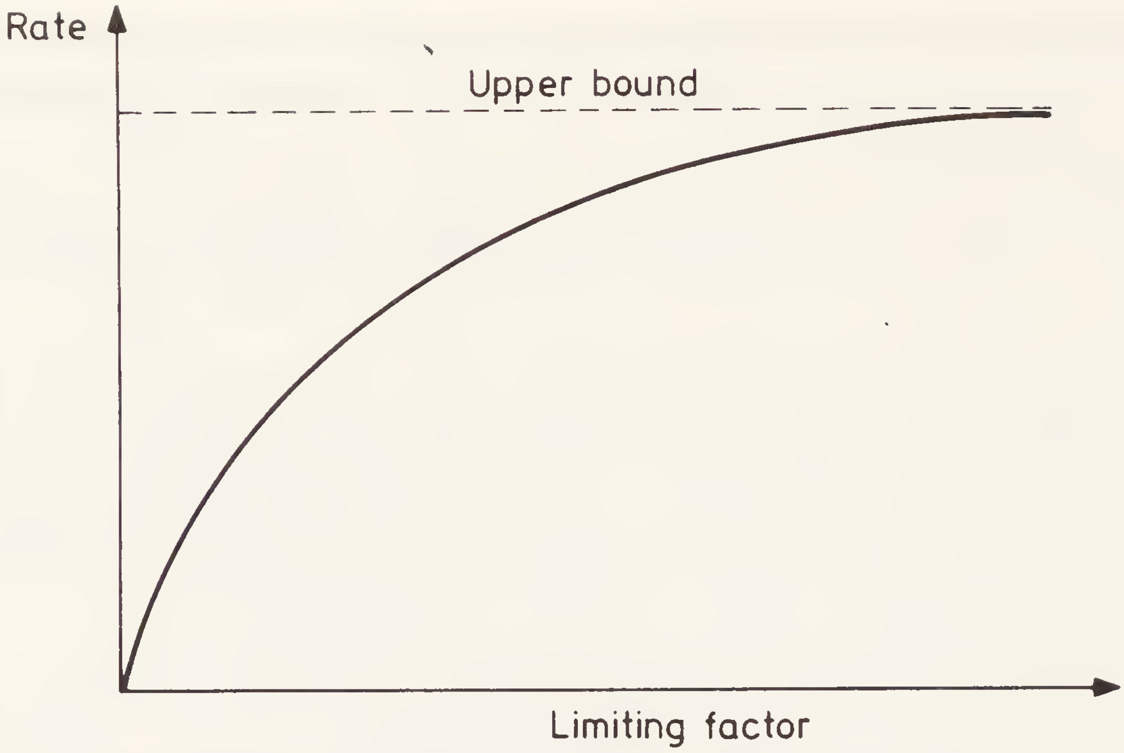


Fig. 5.10. Rate of process vs. intensity (or concentration) of a "limiting factor" such as energy.

body cannot exceed a certain limit precisely because it has a finite amount of energy. Likewise the limited energy resources of the biosphere – in particular of the soil – place a limit on human population. (For further examples and principles see Booij and Wolvekamp 1944.) All such particular statements concerning so-called "limiting factors" can be deduced from formulas regarding the energy of the particular system concerned. We compress them into

POSTULATE 5.5 The rate and the extent of every change in any thing composed of a finite number of things, are bounded.

(The restriction to finite things is due to the need for not excluding the possibility of the ceaseless expansion of the physical universe.)

An immediate consequence of the above postulate is that there is an end to every change. But this matter will be taken up in the next section.

Before closing this section we must ward off a possible misunderstanding concerning universal properties. It is often claimed that information constitutes an even more universal property than energy. However, not all things are information systems: only those constituted by an information source, an information channel, and an information receiver qualify as information systems, i.e. systems capable of generating, transmitting, and receiving information. Secondly, rather than being (relative to every reference frame) an intrinsic property, like energy, information is a mutual property in the sense that it is always an amount of information transmitted by an information carrier: change the channel (e.g. make it less noisy) and the amount of information will change. A cursory look at the basic assumptions and definitions of statistical information theory (e.g. Luce, 1960) will suffice to bear out our contention. Thirdly, while there can be energy exchanges without any information transmission – e.g. in atomic collisions – every transmission of information requires some energy. The fact that information theory pays no attention to the physical basis of an information process, hence to the energy exchanges underlying the transmission of information, does not annihilate that basis. We shall return to this matter in Sec. 4.4.

3. PROCESS

3.1. *Serial Change*: *Types*

Heretofore we have studied change in general. In this section we shall analyze a particular kind of change, namely serial change or process. Not every old set of states or of events constitutes a process. For example an arbitrary collection of events occurring in different things that are comparatively isolated from one another does not constitute a process even if it is ordered in time. For a set of events to constitute a process it must satisfy two conditions:

(i) the events must involve or concern just one thing, however complex, and

(ii) the events must be ordered intrinsically.

In sum a process may be envisaged as a directed tree, in particular a chain, in some event space.

Before laying down general principles we may as well get hold of a stock of typical examples. We shall resort to these examples several times in the course of this work, so we had better display them ostensibly.

(*a*) *Chain*

The simplest process kind is the chain formed by a denumerable set of states, or of events, in a given thing. In this particular case we can formalize the notion of a process as a *sequence*, i.e. a function defined on the set $\mathbb{N}$ of natural numbers. More precisely, we can make

DEFINITION 5.16 Let $\mathbb{F}$ be a state function, and $S(x)$ the corresponding state space, for a thing x relative to a certain reference frame. Then x undergoes a *chain process* iff $\mathbb{F}: \mathbb{N} \to S(x)$, where $\mathbb{N}$ is the set of non-negative integers. In this case the process is represented by

$$\pi(x) = \langle \mathbb{F}(n) | n \in \mathbb{N} \rangle,$$

where $\mathbb{F}(n)$ is the nth *stage* of the process.

A finite segment of such a process is the restriction of the state function to a finite subset of natural numbers. The *length* of such a finite chain is of course the number of stages in it.

Example A population of unstable particles, or organisms, or any other things that come into and go out of being, will undergo a discontinuous process as far as the population number is concerned. If the rate of destruction is balanced by the rate of emergence, a steady state prevails in the given respect.

Discontinuous processes, though common and of great heuristic importance, are not universal. An even when they do take place it is worth while to try to analyze them further instead of taking the component events *en bloc* as was done above. This need not be done in the preliminary stages of a science – e.g. in learning theory at the present moment – but it is an ultimate desideratum. Finally because discontinuous processes do not exhaust the process genus, it makes no sense to ask about the number of events that have occurred during the history of the universe – not even during the last second of its existence.

(*b*) *Continuous process*

An obvious generalization of the denumerable chain is obtained by substituting an arbitrary directed set for the set of natural numbers in the previous definition. More exactly, we propose

DEFINITION 5.17 Let $\mathbb{F}$ be a state function, and $S(x)$ the corresponding state space, for a thing x relative to a certain reference frame. Then x undergoes a *serial change* iff $\mathbb{F}: T \to S(x)$, where T is a directed set [i.e. if $\mathbb{F}$ is a *net*]. In this case the process is represented by

$$\pi(x) = \langle \mathbb{F}(t) | t \in T \rangle.$$

Serial changes are often gapless or continuous, as when the state function is a continuous function of a parameter such as time. We need therefore

DEFINITION 5.18 Let $\pi(x) = \langle \mathbb{F}(t) | t \in T \rangle$ be a serial change in a thing x. Then $\pi(x)$ is *continuous* over the interval $U \subseteq T$ iff U is nondenumerable and $\mathbb{F}$ is a continuous function on U. And the process is *piece-wise continuous* over T iff T equals the union of finitely many nondenumerable sets, and $\mathbb{F}$ is continuous over every one of them.

Example 1 A reversible chemical reaction

$$A \underset{k_-}{\overset{k_+}{\rightleftharpoons}} B$$

in a closed reactor is described by the rate equations with positive coefficients

$$\frac{dA}{dt} = -k_+A + k_-B, \qquad \frac{dB}{dt} = k_+A - k_-B$$

subject to the closure or conservation condition $d(A+B)/dt = 0$. The state space of the whole thing is the $\mathbb{R}^+ \times \mathbb{R}^+$ plane and the evolution of the system is represented by an arc of a continuous curve in this plane.

Example 2 The trajectory of a particle that collides with other particles is broken but continuous even though the particle velocity changes abruptly at each collision. The simplest case is that in which the particle is initially at rest and obtains a constant velocity at $t=0$ upon being pushed by another particle:

$$\dot{x}(t) = \begin{cases} 0 & \text{for} \quad t < 0 \qquad \therefore \quad x(t) = a = \text{const.} \\ 1 & \text{for} \quad t \geqslant 0 \qquad \therefore \quad x(t) = a + t. \end{cases}$$

It is sometimes claimed that continuity is more basic than discontinuity – and at other times that the converse is true. There seems to be no firm ground for either hypothesis, whence we shall adopt neither. We only seem to know that (*a*) both hypotheses have been ontologically and scientifically stimulating; (*b*) continuity hypotheses are empirically irrefutable or nearly so, and (*c*) while some properties (hence some processes) are continuous, others are not – at least according to contemporary science. For these reasons, instead of committing ourselves to either hypothesis we propose the more fertile heuristic

PRINCIPLE 5.4 Try to resolve every discontinuous change into a continuous process, and watch for discontinuities in every seemingly continuous process.

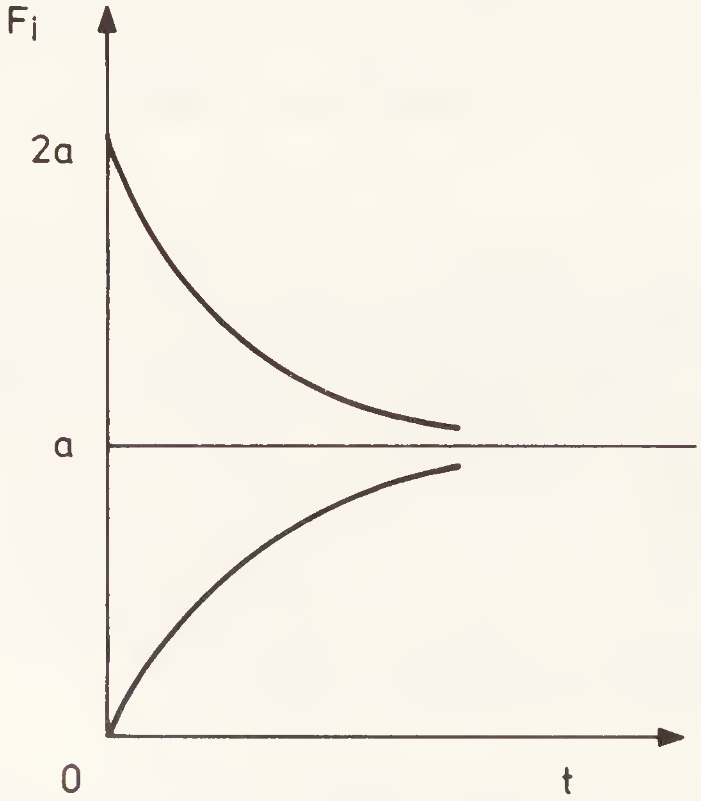
(c) *Path independence*

Whereas the outcome of some processes depends critically upon the precise trajectory in state space, others are *path independent* or nearly so. A particular case of great interest in all fields of inquiry, from physics to sociology, is that of equifinality, or same final state. The precise concept is elucidated by

DEFINITION 5.19 Two serial changes are *equifinal* iff their end states are the same.

Equifinality has been claimed to prove finality or goal-directedness, for it seems as if the system concerned – e.g. an organism – strives to attain a given goal by whatever means it can. However, equifinality is not evidence for teleology unless the existence of a purpose can be proved, for equifinal processes are common among physical things. For example, all closed thermodynamic systems approach a state of equilibrium.

Example Let the instantaneous value of a property F be given by one of the following functions on $\mathbb{R}^+$:



$$F_1(t) = a(1 + e^{-kt})$$
$$F_2(t) = a$$
$$F_3(t) = a(1 - e^{-kt})$$
with $a, k \in R^+$.

Fig. 5.11. Three different ways of reaching the same final state or asymptotic value $X_\infty = a$: starting from a, advancing to it, or declining to it.

(*d*) *Memory*

Suppose the state of a thing depends not only upon a parameter t (the phase of the process) but also upon some preceding state or states. In this case one says that the thing has a *memory* of some of its past states, or that it undergoes a *hereditary* process. Thus the elastic state of a piece of steel, the magnetic state of a magnet, and the genic state of an organism, all depend on their previous histories. In a number of cases such dependence can be expressed by a formula of the type

$$\mathbb{F}(t) = \int_0^t \mathbb{F}(\tau - t) \cdot M(\tau) \cdot d\tau,$$

where M is the memory function. (The process is memoriless iff $M = \delta$, or the Dirac delta.)

The general concept is clarified by

DEFINITION 5.20 Let $\mathbb{F}$ be a state function and $S(x)$ the corresponding state space, for a thing x relative to a certain reference frame. Further, call $t \in T$ the state parameter, where T is a directed set. Then x undergoes a *hereditary process* iff every one of its states $\mathbb{F}(t) \in S(x)$ is a functional of some other states $\mathbb{F}(\tau)$ with $\tau < t$.

There are two important species of hereditary process: (*a*) the memory is kept fairly intact throughout the process, and (*b*) the memory fades away until it disappears practically after a certain time called the *relaxation time*. Systems with weakly interacting components, such as non-ionized gases and extended empires, have short relaxation times. On the other hand organisms have long genetic and ontogenetic relaxation times. For example, malnutrition and deprivation of sociality in the early life of a mammal are separately sufficient to seriously impair some of its abilities for life.

(*e*) *Reversibility and irreversibility*

Some processes, particularly on the microphysical level – and probably only such – are reversible; all known macroprocesses are irreversible. The latter may be regarded as a generalization of a theorem proved in statistical mechanics. A system is called *ergodic* if it eventually scans its entire state space – or, more precisely, if, when left undisturbed (isolated), it occupies successively every state consistent with its total energy.

Until 1913 most authors asserted that, since individual particles move reversibly, every collection of particles must have the reversibility property and, moreover, must be ergodic. That year A. Rosenthal and M. Plancherel proved independently that, if the system satisfies the canonical equations (Ch. 3, Sec. 2.6), then it is not ergodic. That is, they proved that there are no classical-mechanical ergodic (or fully reversible) systems. Hence even if the universe were a huge aggregate of particles (which it is not) it could not retrace all of its steps. There is no eternal return.

Finally we note that in microphysics most, but by no means all, processes are assumed to be *stochastically reversible*, in the sense that the probability of an elementary process and that of its reverse are the same. This allows for momentary deviations from the general trend, i.e. for isolated events in which some states become overpopulated or underpopulated. However, certain microprocesses, notably radioactivity and, in general, transmutation processes, are irreversible.

All this invites

DEFINITION 5.21 Let $\pi(x) = \langle \mathbb{F}(t) | t \in T \rangle$ be a serial process in a thing x. Further, let $\bar{T}$ be the same set as T directed in the opposite way (e.g. $\langle \mathbb{R}, > \rangle$ instead of $\langle \mathbb{R}, < \rangle$). Then $\pi(x)$ is *reversible* iff $\bar{\pi}(x) = \langle \mathbb{F}(t) | t \in \bar{T} \rangle$ is just as possible (i.e. lawful) as $\pi(x)$.

If both $\pi(x)$ and its reverse $\bar{\pi}(x)$ can be assigned probabilities, one can think of defining the *degree of reversibility* of a process as the ratio of its probability and that of its reverse, or $r = \bar{p}/p$. For a fully reversible process, $r = 1$; for a fairly reversible process, $0 \ll r < 1$; for a nearly irreversible process, $0 < r \ll 1$; and for a fully irreversible process, $r = 0$. Such a concept might be useful in a number of fields, particularly in statistical mechanics.

It is common to declare a process to be reversible just in case its laws are invariant under time reversal. This is a mistake, because processes depend on circumstances as well as laws. For example, although the basic laws of radioactive disintegration (chiefly Schrödinger's) are T-invariant, the boundary conditions are such that the probability for a disintegration fragment to reenter the nucleus vanishes.

Finally, note that, while all reversible processes are cyclic, the converse is false. In fact most cycles are not reversible – i.e. they do not leave things as they were. For example, the water cycle and the nitrogen cycle on our planet are not reversible, as they are accompanied by energy degradation (entropy increase).

(*f*) *Random process*

A conspicuous kind of serial process, and one of great philosophical interest, is that of a random process. (For a definition of randomness see Ch. 4, Sec. 6.4.) A random or stochastic process is one described by a state function that is a random "variable", i.e. a function every value of which has a definite probability or probability density. In the simplest case the values s_i of the state function $\mathbb{F}$ form a discrete sequence $\langle s_i | i \in \mathbb{N} \rangle$, such as the times of arrival of customers at a counter. In this case the function f such that $f(s_i) = Pr(\mathbb{F}(t) = s_i) = p_i$, where Pr is a probability measure and p_i the probability that x happens to be in state s_i, is called the *probability distribution* of $\mathbb{F}$. (See e.g. Feller, 1968, Ch. IX.)

Many substantial properties, at all levels, must be represented by random "variables". Hence any real process – in contrast to an idealized model of it – is likely to have at least one random aspect. Therefore, instead of defining the concept of a real process (as is usual in the literature on stochastic processes) we prefer to adopt

DEFINITION 5.22 Let $\mathbb{F} = \langle F_1, F_2, \ldots, F_n \rangle$ be a state function for a thing x undergoing a serial change $\pi(x) = \langle \mathbb{F}(t) | t \in T \rangle$. This process is said to be *random in the ith respect* iff the component F_i of $\mathbb{F}$ is a random variable.

(*g*) *Stability*

A thing is said to be stable if its representative point in state space remains confined within a "small" region. More precisely, we can make

DEFINITION 5.23 Let $G_{\mathbb{L}}(x)$ be the set of lawful transformations of the state space $S_{\mathbb{L}}(x)$ of a thing x. Then x is *stable in the region* $A \subset S_{\mathbb{L}}(x)$ iff, for all $s \in S_{\mathbb{L}}(x)$, $g(s) \in A$ for every g in $G_{\mathbb{L}}(x)$.

In other words a thing is in a state of *dynamical equilibrium* if all the changes it undergoes send it to a fixed subset of its state space. If this subset reduces to a point then the thing is said to be in *static equilibrium* or not to change at all. In other words, a thing x is in static equilibrium at a state $a \in S_{\mathbb{L}}(x)$ iff a is a fixed point under every transformation of the state space into itself – i.e. if $g(a) = a$ for all g in $G_{\mathbb{L}}(x)$. In short, just as fixed subsets of the state space represent dynamical equilibrium, so fixed points represent states of static equilibrium.

States of static equilibrium can only be temporary. On the other hand dynamical equilibrium can be lasting. Figure 5.12 exhibits the states of

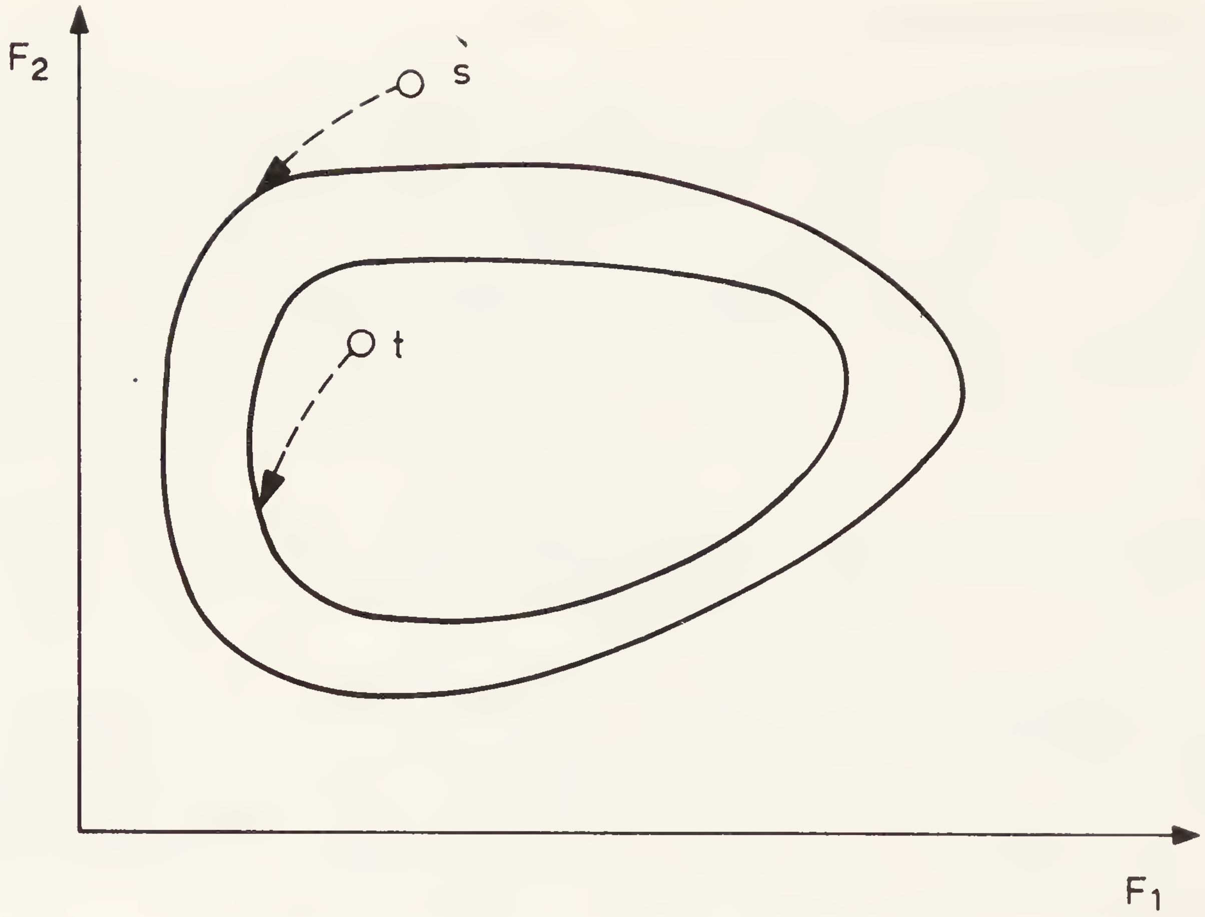


Fig. 5.12. Ecological equilibria are dynamical. If an external perturbation, such as farming or a draught, sends the representative point to s, or to t, the point eventually moves to an equilibrium orbit. (See e.g. Pielou (1969).)

dynamical equilibrium of an ecosystem consisting of two competing populations. Each curve represents a cyclic process – an oscillation in the population numbers. Each point lying on one of the closed curves represents a stable state of the system. If the populations are displaced – e.g. by some external disturbance – to values not too far from equilibrium, they tend spontaneously to return to an equilibrium point.

So much for special types of process. Let us hasten to warn that, in our view, no real process is of a pure type. This is because any (realistic) state function has a number of components of different kinds – some discrete, others continuous, some random, others nonrandom, and so on. In other words, call

$$\pi_k(x) = \langle F_k(t) | t \in T\rangle, \quad \text{with} \quad 1 < k \leq n,$$

the kth *component* of the serial process represented by $\pi(x) = \langle \mathbb{F}(t) | t \in T\rangle$. Then $\pi(x)$ will have at least one continuous component, at least one random component, and so on. In short, pure process types are ideal types: real processes are impure. We defer an exact formulation to the next subsection.

We first turn to some generalities that can be formulated in fairly exact terms.

3.2. *General Concepts and Principles*

In the previous subsection we discussed a few typical kinds of process. In the present one we shall treat processes in general, using some insights gained in the study of special process types. From that study it is apparent that a process may be regarded either as a *sequence of states* or as a *sequence of events*. In the former case we can describe a process in a thing x with the sole help of a state function $\mathbb{F}$ for it: our task boils down to following the wanderings of the tip of $\mathbb{F}$ – a task greatly facilitated by parametrizing the states $\mathbb{F}(t)$ relative to the states of a suitable frame. If on the other hand we construe a process as a sequence of events, then we must be given not only $\mathbb{F}$ but also the entire state space $S_{\mathbb{L}}(x)$ spanned by $\mathbb{F}$ as well as the totality $G_{\mathbb{L}}(x)$ of lawful transformations of the state space into itself. This alone allows us to form the individual events $\langle s, s', g\rangle$ and thereupon the whole event space $E_{\mathbb{L}}(x)$.

Each of these alternative construals has its advantages and disadvantages. The successive states approach is mathematically simpler, if only because the state space is n-dimensional whereas the corresponding event space is $2n$-dimensional. On the other hand the successive events approach, though mathematically clumsier, is philosophically more perspicuous. We can use either approach interchangeably and shall do so. In the present subsection we shall discuss some general principles in terms of events, but shall end up by introducing a notion of history based on the successive states approach.

First the general notions of process:

DEFINITION 5.24 Let $E_g(x)$ be the space of events of type g in thing x, for g in the set $G_{\mathbb{L}}(x)$ of lawful transformations of the state space chosen for x, and let $E(x)$ be the union of all $E_g(x)$'s, i.e. the total event space of x [Definition 5.9]. Further, let $<$ be the precedence relation [Definition 5.10]. Then

(i) a *process of type* g in thing x is a strictly partially ordered set of events of kind g:

$$\pi_g(x) = \langle E_g^*(x), <\rangle \quad \text{with} \quad E_g^*(x) \subseteq E_g(x) \subseteq E_{\mathbb{L}}(x);$$

(ii) a *process* in x is a strictly partially ordered set

$$\pi(x) = \langle E^*(x), <\rangle \quad \text{with} \quad E^*(x) = \bigcup_{g \in G_{\mathbb{L}}(x)} E_g^*(x).$$

Since as a matter of fact every aspect g is related to some other aspect, we may safely surmise that there are no processes of a single type but only idealizations that single out now one feature of the thing of interest, now another. (Recall the end of the last subsection.) That is, we assume

POSTULATE 5.6 Every real process has more than one aspect.

And from the assumed lawfulness of the transformations in $G_{\mathbb{L}}(x)$ it follows that every process involving a given thing satisfies all of the laws possessed by the latter:

COROLLARY 5.9 Every process is lawful.

Note that lawfulness is predicated of processes, i.e. intrinsically ordered sets of events, not of arbitrary sets of events. The latter, particularly if they occur in unrelated things, or in things with weakly coupled components, are not necessarily lawful even though every component is assumed to have a lawful behavior line. This is the ontological ground of the methodological rule employed tacitly by historians intent on discovering general historical trends or even laws – namely to focus on processes of a definite kind occurring in social systems instead of just looking for time series, which are bound to be erratic or lawless (see Braudel, 1969).

Remember however that chaos is not the same as randomness: whereas the former is lawless the latter possesses stochastic (probabilistic) laws such as those of quantum mechanics and genetics. Therefore we adopt an axiom employing Definition 5.22 of a process random in some respect:

POSTULATE 5.7 Every process is random in at least one respect.

Next we distinguish two important members of a process:

DEFINITION 5.25 Let $\pi(x) = \langle E^*(x), <\rangle$ be a process in a thing x. Then

(i) $\pi(x)$ *begins* at an event $b \in E^*(x)$ iff b is a lower bound of the set $E^*(x)$ [i.e. if every other member of $E^*(x)$ follows b];

(ii) $\pi(x)$ *ends* in an event $e \in E^*(x)$ iff e is an upper bound of the set $E^*(x)$ [i.e. if all the other members of $E^*(x)$ precede e].

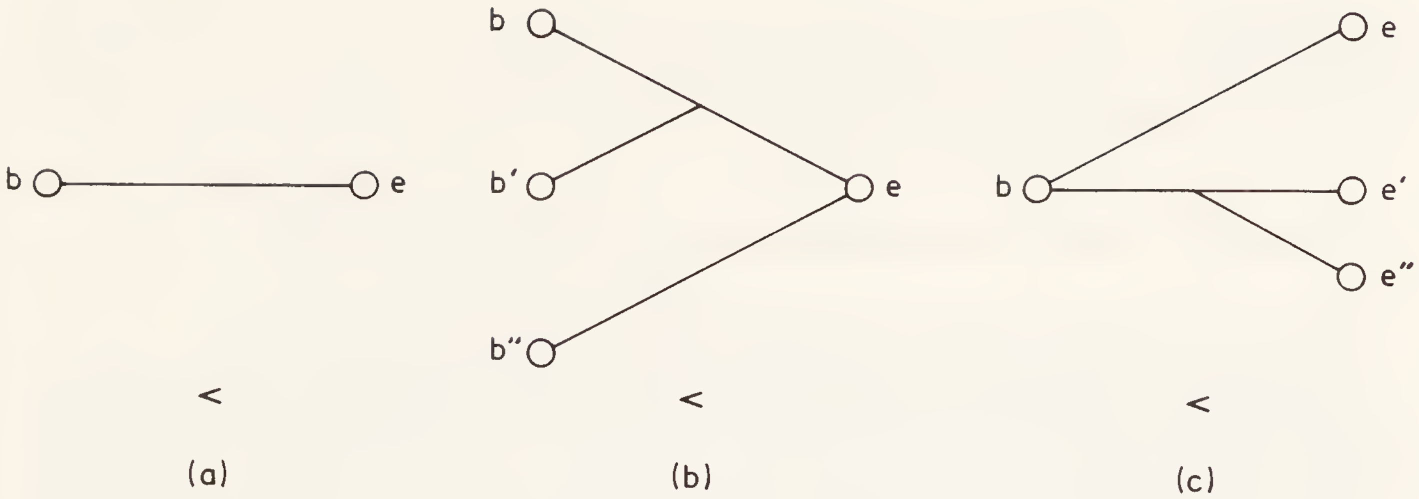


Fig. 5.13. Three kinds of process as to beginning and end. (*a*) Chain or linear process. (*b*) Converging or branching-in process. (*c*) Diverging or branching-out process.

A process may have no single beginner or single stop: see Figure 5.13. In other words, there may be multiple beginnings or multiple ends. We assume that every single process is bounded both below and above:

POSTULATE 5.8 Every process has beginning(s) and end(s). [I.e., if $\pi(x)$ is a process going on in a thing x, then $\pi(x)$ has at least one beginner and at least one end point.]

This hypothesis does not entail that there are absolute beginnings (i.e. processes starting out of the blue) and dead ends (i.e. processes that, on ending, do not start any other processes). Indeed the postulate is compatible either with a creationist and eschatologist cosmology or with the assumption that every process in a given thing, though bounded both below and above, is preceded and followed by some other processes either in the same thing or in other things. We adopt the latter hypothesis because it is the one supported by science. But first we need

DEFINITION 5.26 Let $\pi(x) = \langle E^*(x), \prec \rangle$ and $\pi'(x) = \langle E'^*(x), \prec \rangle$ be two processes in a thing x, such that $E^*(x) \cap E'^*(x) = \varnothing$. Then $\pi(x)$ *precedes* $\pi'(x)$ iff every event in the former precedes every event in the latter. Symbol: $\pi(x) \prec \pi'(x)$.

We are now ready for

POSTULATE 5.9 Every process is preceded and followed by other processes. More precisely, every thing x has parts y and z, not necessarily different from x, each of them equipped with its own event space, such that

$$\pi(y) < \pi'(x) < \pi''(z).$$

Since the world is the aggregation of all things, it follows

COROLLARY 5.10 The world has neither beginning nor end.

Remark 1 Like many another axiom of our ontology, the previous one is a sort of generalization of all we know. An ancient paradigm is of course the birth-growth-final disintegration process every metazoan undergoes. A simpler and more modern paradigm is that of the photon emitted by a star, travelling through interstellar space, and ending up trapped by a molecule or turning into a pair of electrons of opposite charges. *Remark 2* At one point some cosmologists wishing to save the steady state cosmology introduced the *ad hoc* hypothesis of the spontaneous creation of hydrogen atoms in empty space. This conjecture was not accepted because it clashes with the conservation laws (Bunge, 1962) – and anyway the steady state theory, which it tried to save, was eventually refuted by astronomical evidence. *Remark 3* Corollary 5.10 was first formulated by the Presocratic philosophers and was adopted by Aristotle. It would seem to contradict the "big bang" hypothesis of the "origin" of the universe, nowadays almost universally favored. However, this cosmological conjecture – which explains the expansion of the universe – does not illustrate creation *ex nihilo*, for it presupposes that the universe had been in existence – in a highly condensed state – before the expansion started. Nor does the old "thermal death" hypothesis (nowadays almost forgotten) illustrate absolute annihilation: all it asserts is the end of macrophysical motion once a closed system has attained thermal equilibrium.

So much for our general concepts and principles concerning process. Let us now adopt the successive states approach and introduce one more key notion that will prove indispensable in the sequel, namely that of history.

Recall that the states of a thing x can be coordinatized relative to the states $t \in S(f)$ of a reference frame f, by defining the state function as a mapping of $S(f)$ into $S(x)$, and by construing $S(f)$ as a four-dimensional grid conceptualizing a physical reference frame (Sec. 2.5). As the reference state t "moves" so does the tip $\mathbb{F}(t)$ of the thing state function. This invites

DEFINITION 5.27 Let $\mathbb{F}$ be a state function for a thing x relative to a reference frame f with states t in $S(f)$, and let the latter be coordinatized by a certain function $k: \mathbb{R}^4 \to S(f)$. Then the *history of x relative to f* is the set of ordered pairs

$$h(x) = \{\langle t, \mathbb{F}(t)\rangle \mid t \in S(f)\}.$$

This is an elucidation of the concept of a life line, behavior line, or trajectory. The history of a thing is regarded as the succession of its states but, instead of being represented by a line in the n-dimensional state space spanned by $\mathbb{F}$, it is represented as a curve in the $(n+4)$-dimensional space $\mathbb{R}^4 \times S_{\mathbb{L}}(x)$. (The dimensionality of this space can be reduced by 3 if we are interested only in the time coordinate, i.e. if t is construed as an instant of time rather than as a point in the spacetime region attached to the given frame.)

The above definition covers the most usual notions of history employed in the sciences. In particular it fits in with the notion of history occurring in one of the most fundamental theories in physics, namely quantum mechanics. Indeed in this theory the evolution operator for a thing is defined as

$$S(t_0, t) = \exp\left(-\frac{i}{\hbar}\int_{t_0}^{t} du\, H\right),$$

where H is the hamiltonian of the thing. The history of the latter is then

$$h(x) = \{\langle t, S(t_0, t)\psi(t_0)\rangle \mid u \in [t_0, t] \ \&\ \psi \in \text{Hilbert space of } x\}.$$

The broad concept of history adopted above may puzzle and even infuriate those scholars in the humanities who still regard man as supernatural. Until about the mid 19th century it had been tacitly assumed that all history is human history. Thus Hegel, for all his dynamicism, denied that nature has a history. That view changed with the revolutionary discoveries in geology and biology, to the point that Engels declared each science to be the history of something (Engels, 1878). Since then it is usually taken for granted that ***every thing has a history*** – i.e. that, for every thing x, $h(x)$ is bigger than a singleton.

Let us now apply the concept of history to a characterization of the notions of action and interaction.

4. ACTION AND REACTION

4.1. *Induced Change*

So far we have been concerned with change without regard for its sources. Now, a thing may change either by itself or under the influence of some other thing, i.e. spontaneously or by induction. We must therefore study the concepts of spontaneity and induction. To do this we shall avail ourselves of the concept of history (Definition 5.27 in Sec. 3.2) and of Leibniz' law (Ch. 2, Sec. 3.2).

Two different things of the same kind may be described with the help of the same functional schema $\langle M, \mathbb{F}\rangle$, hence of the same state space, though of course adding individuating circumstances in each case. In other words, the states of different kindred things may be represented in one and the same state space. Also: every state space, far from being owned by a single thing, may be assigned a reference class constituted by a more or less numerous equivalence class of things. Now, Leibniz' law (Postulate 2.5 or Corollary 2.1) states that different things have different properties. Hence even if a given state function describes adequately two things, they cannot occupy exactly the same states – for, if they did, they would be one. Consequently we have the following rephrasing of Leibniz' law:

COROLLARY 5.11 Different things have different histories.

Surely a thing can occupy now exactly the same state another thing was in a moment ago, but no two things can occupy exactly the same state at the same time – or, more generally, for the same value of the change parameter. (See Figure 5.14.)

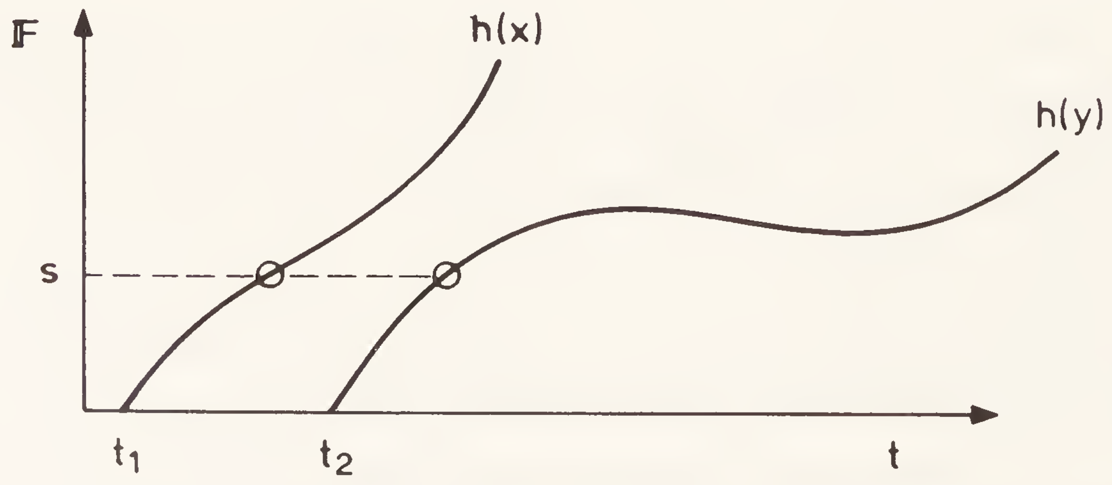


Fig. 5.14. Different things can occupy the same state provided they do it at different times.

Corollary 5.11 is our present version of Leibniz' *principium individuationis* and it supplies the clue to what we are after. Indeed, if a thing is under the action of another thing, then its history must differ from the history when free from such action. (See Figure 5.15.) (This is what scientific experiment, as different from observation, is all about, namely the deliberate switching on and off of actions or influences to compare the forced behavior of a thing to its free behavior and thus to assess the importance of the variable that is being manipulated.)

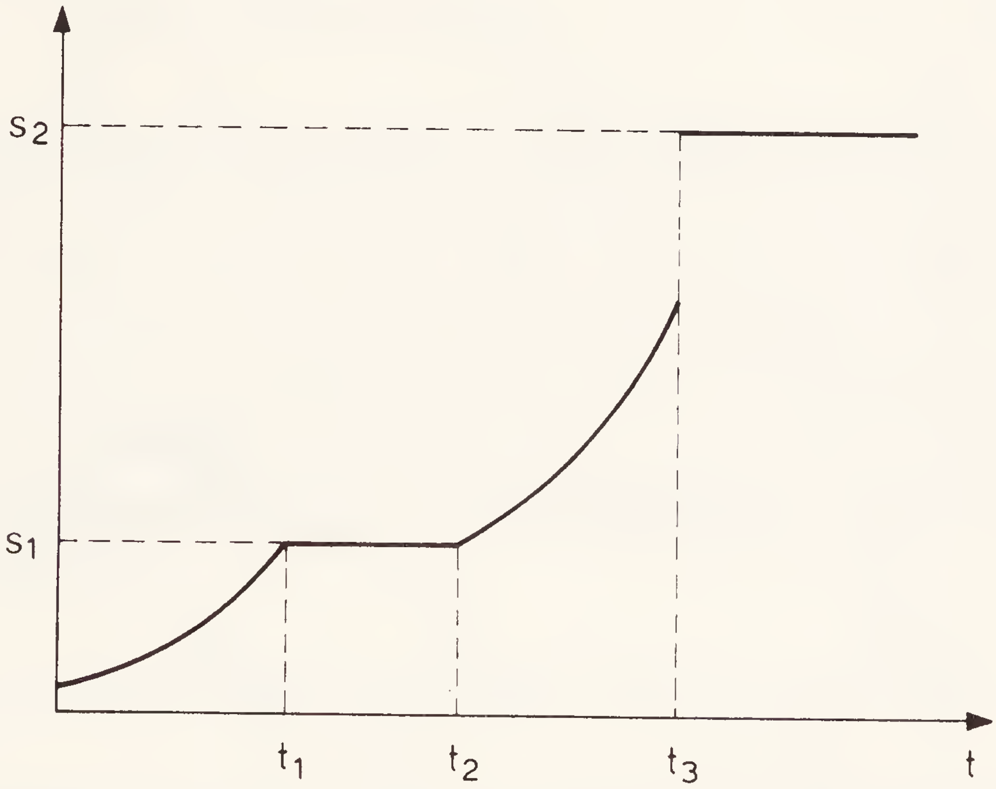


Fig. 5.15. Imaginary example of an action. The agent, by acting during the $[t_1, t_2]$ interval, forces the patient to remain in state s_1. The external action resumes at time t_3, forcing the patient to jump to state s_2.

In the absence of any actions of the agent x upon the patient y, the history of their physical sum or juxtaposition $x \dotplus y$ equals the union of their individual behavior lines. In other words, we have

DEFINITION 5.28 Thing x changes by itself, or *spontaneously*, iff $h(x)$ is not a singleton and, for any other thing y,

$$h(x \dotplus y) = h(x) \cup h(y).$$

Clearly, if one of the things acts upon the other, then while the behavior line of the agent remains unaffected, that of the patient becomes, as it were, the history of a different thing, namely the thing-acted-on-by-the-agent. (See Figure 5.16.) More exactly, we make

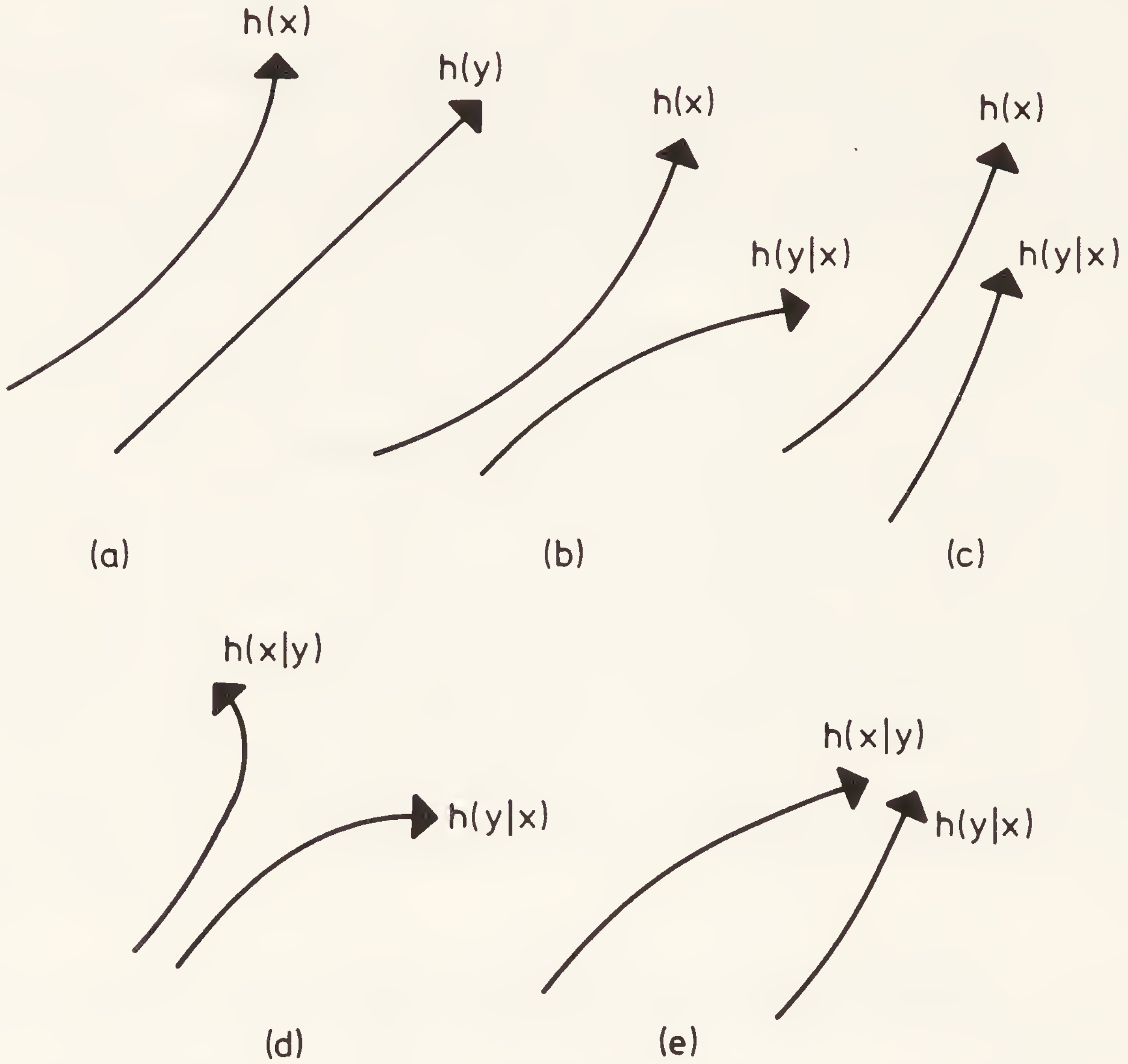


Fig. 5.16. Induced change. (*a*) No action: spontaneous change of both x and y. (*b*) Pushing action of x on y. (*c*) Pulling action of x on y. (*d*) Pushing interaction. (*e*) Pulling interaction.

DEFINITION 5.29 Let x and y be two different things with state functions $\mathbb{F}$ and $\mathbb{G}$ respectively relative to a common reference frame f, and let

$$h(x)=\{\langle t, \mathbb{F}(t)\rangle | t \in S(f)\}, \qquad h(y)=\{\langle t, \mathbb{G}(t)\rangle | t \in S(f)\}$$

be their respective histories. Further, let $\mathbb{H}=g(\mathbb{F}, \mathbb{G}) \neq \mathbb{G}$ be a third state function, depending on both $\mathbb{F}$ and $\mathbb{G}$, and call

$$h(y|x)=\{\langle t, \mathbb{H}(t)\rangle | t \in S(f)\}$$

the corresponding history. Then x *acts on* y, or $x \rhd y$ for short, iff, for some state function $\mathbb{H}$ determining the trajectory $h(y|x)$, $h(y|x) \neq h(y)$.

We may now rephrase Definition 5.28 of spontaneity: A thing *changes spontaneously* iff it changes and if there is no other thing that acts upon it. And, from the hypothesis of Definition 5.29 (in particular, that $x \neq y$), it follows that *nothing acts upon itself.*

The definition of reciprocal or mutual action is now obvious:

DEFINITION 5.30 Two different things x and y *interact* iff each acts upon the other. In symbols:

$$x \rhd\!\!\lhd y =_{df} x \rhd y \;\&\; y \rhd x.$$

Experimental scientists assume tacitly (a) that all things are studiable, (b) that there are no totally isolated things – i.e. that every thing acts on, or is acted upon by, some other thing, and (c) such actions are the basis of empirical knowledge. Indeed if a presumed thing does not act on any of our observation means then we cannot observe it; and if some putative thing is not acted upon by any of our experimental means then we cannot experiment on it. In either case we won't bother with "it". In short, the following axiom belongs to the ontology inherent in science, so we adopt it:

POSTULATE 5.10 Every thing acts on, and is acted upon by, other things.

This assumption should not be mistaken for the far stronger, and false, hypothesis that any two things interact – let alone with the same intensity. We do not react back to the light activity of the sun, even less to that of an extinct star. In other words, the universal interaction principle, first suggested by the Stoics and much favored during the acme of Newton's gravitation theory, is false.

The effects of an interaction may be lasting, even in the case of microthings. Thus if two microthings interact for a while and then cease to do so – e.g. by becoming widely separated in space – they will continue to be correlated. Hence observing one of them may yield information about the other – just as observing the behavior of a divorced person may shed light on his/her former mate. In the jargon of quantum mechanics, such things are said to be *non-separable*, or to exhibit an EPR correlation. (This is what the Einstein–Podolsky–Rosen paradox seems to be all about. More in Sec. 4.2.)

We need a qualitative concept of the size of an action or an interaction. This is given by the amount of the distortion effected by the agent(s) on the behavior line of the patient. In other words, we make

DEFINITION 5.31 Let x and y be two things, and let $h(x)$ be the history of x, and $h(x|y)$ the history of x when acted upon by y, and similarly for x. Then

(i) the *total action* or *effect* of x on y equals the difference between the forced trajectory and the free trajectory of y:

$$A(x, y) = h(y|x) \cap \overline{h(y)};$$

(ii) the *total reaction* of y upon x is the difference between the free trajectory and the forced trajectory of x:

$$A(y, x) = h(x|y) \cap \overline{h(x)};$$

(iii) the *total interaction* between x and y is the union of action and reaction:

$$I(x, y) = A(x, y) \cup A(y, x).$$

If $A(x, y) = \varnothing$ we can say that x exerts the *null action* on y.

Since the action of x on y is a set of events in y, this set cannot be the same as the reaction of y on x, which is a set of events in x. Therefore we have

COROLLARY 5.12 No action equals the corresponding reaction. I.e., for any things x and y, $A(x, y) \neq A(y, x)$.

This result is not inconsistent with the third law of Newtonian particle mechanics. Indeed this law does not state that actions and reactions are the same but that their strengths (the corresponding forces, or else torques) are equal. Anyway the law does not hold in other theories, such as electrodynamics.

Now, actions may exist in certain respects but not in others. That is, the action of thing x upon thing y may affect only some of the components of the state function of y. Hence we may analyze the total action into partial actions, namely thus:

$$A(x, y) = \bigcup_{i=1}^{m} A_i(x, y),$$

where m is the number of mutually independent actions that x exerts on y. Similarly for the reaction of y on x and for their interaction.

This distinction allows us to supplement Postulate 5.8 with

POSTULATE 5.11 Every thing changes spontaneously in some respects, inductively in others.

Finally, the concept of action permits us to elucidate the important distinction made in Ch. 2, Sec. 5.1, between a relation that affects its relata and one that does not. (See Figure 5.17.) The difference is rendered precise by

DEFINITION 5.32 Two different things are *bonded* (or *linked* or *coupled*) together iff at least one of them acts upon the other. In symbols: If x and y are things, then

$$Bxy =_{df} x \rhd y \quad \text{or} \quad y \rhd x.$$

Note that we are not including the condition that the bond be attractive. Two enemies are just as bonded as two friends.

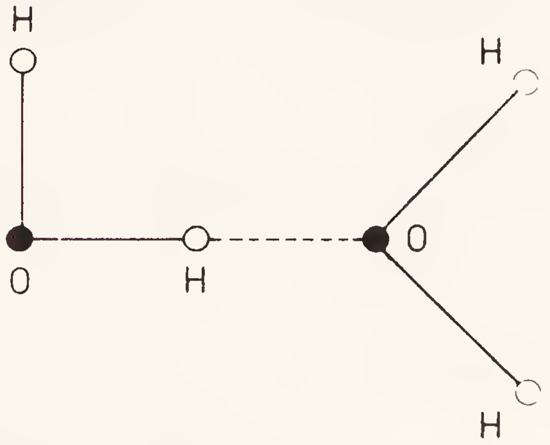


Fig. 5.17. Three different relations between two water molecules. (a) The valence bonds between the atomic components of each molecule (full lines); (b) the hydrogen bond between both molecules (dotted line); (c) the relation of being to the left (or its converse). Both (a) and (b) are bonds, links, or couplings, whereas (c) is a nonbonding relation.

The collection of bonds among members of an arbitrary set T of things deserves a special name:

DEFINITION 5.33 The *bondage* of a set T of things is the set $\mathbb{B}_T$ of all bonds among members of T.

In principle $\mathbb{B}_T$ includes couplings of all n-arities – dyadic, triadic, and so on. In practice the dyadic or binary bonds are by far the most important. If only binary bonds are taken into account, the above definition reduces to

$$\mathbb{B}_T = \{B_i \in \text{Binary relations on } T \mid (\exists x)(\exists y) \\ (x, y \in T \;\&\; B_i xy \;\&\; 1 \leqslant i \leqslant m\},$$

where $m \geqslant 1$ is the number of kinds of binary bond [or of action] holding among members of T.

It will be recalled that we assumed that there are finitely many general substantial properties (Postulate 2.3). Therefore we have

COROLLARY 5.13 The bondage of any set of things is finite.

The class of all nonbonding (or "mere") relations among T's will be designated by $\bar{\mathbb{B}}_T$. And the union of $\mathbb{B}_T$ and $\bar{\mathbb{B}}_T$ equals of course the set of all relations on T. This concept will become crucial in our treatment of the concept of a concrete system in Vol. 4, Ch. 7, so it invites

DEFINITION 5.34 Let T be a set of things and $[T]$ its aggregation [physical sum]. Then the *structure* of $[T]$ equals the totality of relations, both bonding and nonbonding, on T:

$$\mathcal{S}([T]) = \mathbb{B}_T \cup \bar{\mathbb{B}}_T.$$

We shall use the concept of a bond to characterize systems and reference frames.

4.2. *Aggregates and Systems*

We shall now employ the concepts of bond and history to clarify the distinction between a heap or aggregate, on the one hand, and a whole or system, on the other. Obvious examples of systems are atoms and molecules. On the other hand the nucleons and electrons composing an atom are not systems since they have no separable components. Another example of a nonsystem is the aggregate formed by a thing and a reference frame. Since reference frames are supposed to record without either influencing or being influenced, the history of a thing-frame compound must reflect that independence. How this is done will be seen in Sec. 4.3 once we have properly introduced the notion of a system.

An analysis of the histories of the members of a set of things should help us discover whether or not the aggregation of the set constitutes a system, i.e. whether or not its components are bonded. Indeed, the history of an aggregate of noninteracting things is uniquely determined by the histories of such components: it is just their union. Not so in the case of a system: here the history of every component is determined at least partly by the states other components are in, so that the history of the whole does not equal the sum of the individual histories. An example or two will help bring this point home.

Think of the histories of a moth and a candle before and after the former spirals into the latter. If a more distinguished example is needed, think of a system composed of two electrons close enough to interact appreciably, as is the case with those of a helium atom. This system is described by a Schrödinger equation (classically by two coupled equations of motion) jointly with the Pauli exclusion principle. The latter selects those state functions that are odd (antisymmetric) in the coordinates of the electrons. (That is, the additional law is: $\psi(x, y) = -\psi(y, x)$, where x and y are the position coordinates of the electrons. It follows that $\psi(x, x) = 0$, i.e. the two entities cannot coexist at the same point – unless they have different spins.) The Pauli principle expresses a global or systemic property, one that the individual components fail to possess, whence it cannot be represented in the partial state spaces. Therefore the construction of the state space of the system must proceed from scratch rather than on the sole basis of the state spaces of the individual electrons. (Fortunately the Pauli principle does not hold for the entire universe but only for certain parts of it, namely the subsystems formed by half integral spin components. Otherwise there would exist no fairly isolated systems in reality. (Cf. Margenau (1966).) What holds for state spaces holds a fortiori for histories.

We summarize and generalize the preceding remarks into

DEFINITION 5.35 Let X be a thing composed of parts X_i for $1 \leqslant i \leqslant n$. Then X is an *aggregate* (or *conglomerate* or *heap*) iff, in every representation of X (i.e. for every choice of state function), its history $h(X)$ equals the union of the partial histories $h(X_i)$. Otherwise X is a *system*.

The cautious phrase 'in every representation', occurring in the preceding convention, is due to this. It is often possible to choose a representation of a thing wherein the mutual actions of its components are "absorbed" in the latter. In other words one can often model the real system as a conceptual aggregate of noninteracting components. But this trick does not work for every representation.

Recalling the Definitions 5.31 of action and 5.33 of bondage, we obtain

COROLLARY 5.14 Let X be a thing with composition $\mathscr{C}(X)$. Then X is a *system* iff the bondage of $\mathscr{C}(X)$ is not empty, i.e. $\mathbb{B}_{\mathscr{C}(X)} \neq \varnothing$.

Now, the universe or world is that thing which is composed of all things (Postulate 3.3); and by Postulate 5.10 every thing – hence every component of the world – is bonded to some other thing. Hence

COROLLARY 5.15 The world or universe is a system.

By Definition 5.35 it follows in turn that the history of the world is not equal to the union of the histories of its parts.

What happens if a system breaks down, e.g. if its components fly apart and cease to interact? It would seem that, when this happens, the history of the aggregate does equal the union of the histories of the former components, which are now free from one another. This is not necessarily so. In fact the famous Einstein–Podolsky–Rosen paradox of quantum mechanics may be interpreted as follows. If two quantal things interact at any given time then, even after they have ceased interacting, they are still correlated in the sense that the history of each is affected by what happens to the other. (Equivalently: the state vector of the whole does not equal the direct product of the state vectors of the parts.) Thus interaction at some time is sufficient to induce a correlation that is reflected in the nondecomposability of the total state space, and a fortiori in the nonseparability of the partial histories. If this interpretation is correct, then the above definitions may be kept but a new assumption must be added: *Once a* (*micro*)*system always a* (*micro*)*system*. However, this is still debatable, so we shall not incorporate it into our system.

4.3. *Reference Frame*

Another important application of the notion of an aggregate is to the characterization of a reference frame. This concept is of paramount importance not only in physics but also in our ontology, for in real cases states are frame dependent. (The paradigm is of course the state of motion of a body: its track in state space is frame dependent. So much so that it won't move at all relative to a frame rigidly attached to itself.)

One condition for a thing to qualify as a reference frame for another is that the two not influence each other, so that their respective states be neatly separated. (To indulge in anthropomorphism: a reference frame must be an impartial spectator. But this is just a didactic prop: we must not mistake frames for observers.) Another condition is that the states of the reference frame for a given thing be utilizable for parametrizing the states of the thing of interest, in the sense that, for every state t of the reference frame, the thing is in a given state $s = \mathbb{F}(t)$, where $\mathbb{F}$ is the n-component state function for the thing. (See Figure 5.18) These two conditions – independence and parametrizability – will be taken to be jointly necessary and sufficient to characterize the general concept of a reference frame. (Each scientific theory may define its own specific concept of a reference frame by adding its own additional conditions – usually that certain laws be satisfied relative to the frame. See Bunge (1967b)). In sum, we make

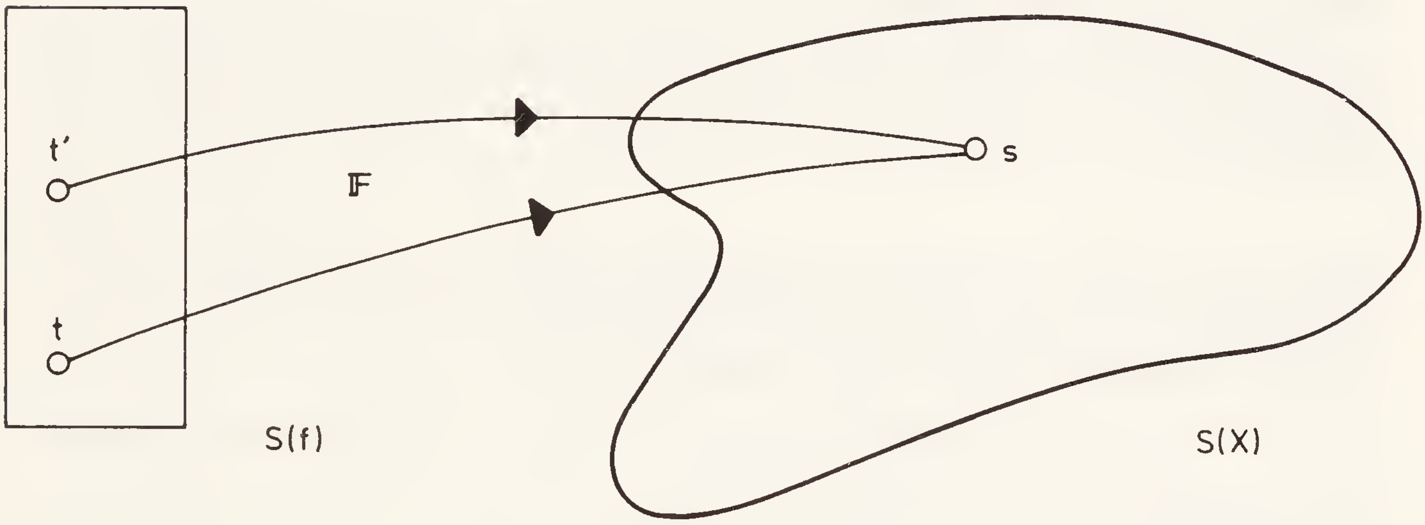


Fig. 5.18. The mapping $\mathbb{F}: S(f) \to S(X)$ of reference states onto thing states. $\mathbb{F}$ need not be 1–1: two different reference states may correspond to a single thing state, as in periodic motion. But $\mathbb{F}$ must be onto: each thing state must be paired off to at least one reference state.

DEFINITION 5.36 Let X be a thing represented by the functional schema $X_m = \langle M, \mathbb{F} \rangle$ and let f be another thing with states in $S(f)$. Then f is a *reference frame* for X iff

(i) $f \dotplus X$ is an aggregate, so that $h(f \dotplus X) = h(f) \cup h(X)$, and

(ii) the domain of the state function of X equals the state space of f, i.e. $\mathbb{F}: S(f) \to S(X)$, whence $M = S(f)$.

Remark 1 We have defined a reference frame as a concrete thing of a certain kind. Because every reference frame can be *represented* by a coordinate patch in a manifold, the misconception has spread that reference frames are identical with their conceptual representatives. That this is mistaken is realized upon recalling that frames are assigned substantial properties, such as velocities, which would be impossible if they were concepts. *Remark 2* A reference frame need not be a perfectly rigid body. This is so firstly because there is no such thing as a rigid body in the real world, secondly because a frame must have moveable parts if it is to function as a (natural or artificial) clock. A frame can be a field (Dehnen, 1970).

The concept of a reference frame allows us to elucidate the notion of a thing's being in a given state relative to some frame. In fact we can introduce

DEFINITION 5.37 Let f be a reference frame for a thing X represented by the functional schema $X_m = \langle S(f), \mathbb{F}\rangle$. Then

(i) the *state of X at $t \in S(f)$ relative to f* is

$$s = \mathbb{F}(t) = \langle F_i(t) | 1 \leqslant i \leqslant n\rangle;$$

(ii) the *state space of X relative to f* is

$$S(X) = \{s = \mathbb{F}(t) | t \in S(f)\} = \mathbb{F}(S(f)).$$

Because there is an unlimited number of reference frames, and the choice of functional schema to represent a thing is partly conventional, states can be parametrized in a number of ways. Some such representations are equivalent – as when the frames are inertial in a certain sense – while others are not. This does not entail that actually states are absolute (frame independent) and that we should blame our own cognitive limitations for the relativity of states. Except in the highly artificial case when all the components of a state function are assumed to be invariant under all possible frame substitutions, *the states themselves are relative.* This is enough to regard with suspicion all the philosophical theories that employ absolute state descriptions such as "Thing X is in state s" or, worse, "The state of the world at a given instant is such and such". (Possible worlds metaphysics and inductive logic abound with such curiosities.) But relativity must not be mistaken for subjectivity: that every state of a thing is relative to a frame does not entail that it is the private state of a subject – unless of course one manages to confuse thing states with reference states or states of a standard of reference, and frames with subjects or observers. (For criticisms of such confusions see Bunge, 1967b, 1973b.)

So much for technicalities. Let us now turn to a few rather nontechnical yet important matters of principle.

5. PANTA RHEI

5.1. *Fact*

Having now at hand exact definitions of a state (Ch. 3, Sec. 2) and of a change of state, or event (Sec. 2), we can make full sense of Definition 4.3 of a fact. We can now say that *a (real) fact is either the being of a thing in a given state, or an event occurring in a thing.* The following comments spring to mind.

Firstly, the above definition of a fact applies also the case of the conversion of one thing (e.g. a caterpillar) into another (e.g. a butterfly), provided we construe both things as a single thing in two successive stages of development. In fact, as suggested in Sec. 1.2, all we have to do is to feign that we are dealing with one and the same thing whose representative point moves now in one state space (e.g. the caterpillar state space), now in another (e.g. the butterfly state space), the two state spaces being correctly assembled.

Secondly, constructs – such as concepts and propositions – do not qualify as facts since they are neither states of things nor changes in things. Hence it is mistaken to call a true factual proposition *a fact.* At most, a proposition might be regarded as a certain equivalence class of mental states – but a class is a concept not a thing. Hence it is mistaken to identify the two.

Thirdly, our construal of facts is objective. We do not define a fact as "anything that may be observed" because (*a*) microphysics, megaphysics and the physiology of perception teach us that most facts remain beyond human observation (though not therefore beyond comprehension), and (*b*) not all observations are reliable: some correspond to no external facts, whence reports on them, unless critical, are false.

Fourthly, the same applies, *mutatis mutandis,* to the semantic definition of a fact as "the kind of thing that makes a proposition true or false" (Russell, 1918, in Marsh, Ed., 1956, p. 183). And this for the following reasons: (*a*) unless one wishes to construct a subjectivist ontology (as was the case with Russell (1914), Whitehead (1919), Carnap (1928), and Goodman (1951), one will refrain from defining ontological concepts in semantical, epistemological, or psychological terms; (*b*) most factual propositions are at best partially true rather than wholly true; (*c*) the semantic definition of a fact commits one to asserting that there are negative facts (those verifying negative statements) and general facts (those confirming general statements). The weird metaphysics of negative and general facts is avoided if, contrary to the early Wittgenstein and the Russell of logical atomism, one drops the assumption that propositions and facts stand in a 1–1 correspondence. The correspondence cannot be such because all facts are "positive" and singular. (Recall that a fact is either the being of a thing in a state, or the occurrence of a change of state of a thing.) The nonoccupancy of a state and the nonoccurrence of an event are not facts. Thus not having been inoculated against smallpox is not an event, hence it does not count as a cause of smallpox. And since we do not countenance negative events, we avoid talking of "logical events" such as "*e* or not-*e*", and so we do not have to worry about their causes. We take this to be a distinctive advantage over the probabilistic theory of causality (Suppes, 1970), where any element of a probability space qualifies as an event and where the causal relation is assumed to hold among possible events, even negative ones, rather than among actual events. In short it is false that whatever makes a proposition true or false counts as a fact.

Fifthly and lastly, facts, whether possible or actual, are objective but – pace Wittgenstein (1922) – they do not constitute the world. The world is not the totality of facts but the totality of things, i.e. concrete individuals endowed with all their properties, some of which consist in their being able to change in definite (though possibly randomly) lawful ways. This is how both physical cosmology and our ontology construe the term 'world'. Wittgenstein's construal requires either dispensing with things altogether – an awkward situation for science – or attempting to define them in terms of facts – which attempt has not even been broached. More on this in Sec. 6.

5.2. *Dynamicism*

Ours is a world of things – but of changing things not quiescent ones. Indeed, according to Postulate 5.11 in Sec. 4.1, every thing has a nontrivial history. In other words *every thing changes*. This is our version of the Heraclitean *panta rhei.* Note that we are not yet taking a stand concerning the stronger ontological hypothesis that every thing changes *incessantly*: we do not have yet a suitable concept of time. For Postulate 5.11 to hold it is necessary and sufficient that every thing changes at least once – i.e. that the history of every thing consists of at least two states or, equivalently, of one event.

We take change – not its concept – for granted rather than either deny or justify it. What has got to be justified in each case is not change but any assumption, or pretence, to the effect that a certain thing fails to change in some respect. We do this kind of thing all the time, by saying that certain changes, though real, are small, or that they consist in nothing but relative motion, etc. Moreover we explain permanence (in certain respects) in terms of insulation, or isolation, or a balance of forces, or what not. Statics is a particular case of dynamics, electrostatics of electrodynamics, and so on. The most powerful and deep theories of contemporary science are theories of change not theories of being: permanence is a particular, exceptional, temporary case of change. To deny that all things are restless, to deny that universal change is objective – as Weyl (1949), Costa de Beauregard (1963) and Grünbaum (1967) – have done is an exercise in sophistry.

The dynamicist view inherent in contemporary science, and adopted by our ontology, contrasts not only with Parmenidean staticism but also with the Aristotelian and Thomist hypothesis that rest, rather than motion, is the "natural" state of things. Our version of dynamicism goes also one step beyond Democritus', for whom atoms themselves were subject to no changes. Both high and low energy physics suggest that not even the "ultimate" components of things – "elementary particles" and fields – are quiescent. There is no inert matter, no perpetually unchanging stuff that can be disturbed only by external influences but is otherwise passive. As Bolzano saw, such external influences and the resulting relative motions could not be explained "if no change could take place in the interior of the simple substances themselves"; hence "the capacity for change through mutual influence" must be attributed to all substances, whether complex or simple (Bolzano, 1851, §§ 50, 51). The "immutable pole" exalted by some poets and philosophers is not a thing or a set of things but the set of basic objective laws.

Our view is definitely dynamicist but not excessively so; we assume that every thing changes in some respect or other but refrain from asserting that it changes in every respect, let alone incessantly. Science does not seem to support such an exaggeration: it shows that some traits remain invariant throughout certain changes. (The constants of the motion of a physical system are among such invariants.) Moreover change in some respects is possible only on account of permanence in others. (Thus, life processes are impossible unless homeostasis is maintained.) Likewise whatever is permanent (invariant, unchanging) is so relative to (under) some specific set (groupoid, semigroup, or group) of transformations. Consequently the invariants of any one such structure may not be the same as those of a different set of transformations.

What holds for every single thing – namely that it is changeable – is true of the thing composed of all things, namely the universe. That is, the world as a whole is in flux – though not as a single well integrated system or cosmos. The universal flux is a huge bundle of processes some of which intertwine (influence each other) while most probably do not. Hence it is mistaken to speak of "the world series of events" – as if it were something like an oversized world baseball series. A fortiori it is mistaken to affirm or to deny that such a series has a beginning or an end. There just *is* no world series of events. Because the precedence relation among events is local (bound to some reference frame or other), the totality of events has no overall order. All we can assert – and have in fact asserted in Postulate 5.9 – is that every single process is preceded and succeeded by other processes. Hence although the world as a whole is not engaged in a relay race, every bit of it does participate in some local relay race or other.

5.3. *Interconnectedness*

Postulate 5.10 in Sec. 4.1 states that every thing influences, or is influenced by, some other things. In other words, nothing – except for the universe as a whole – is totally isolated. If this is so then the world is an interconnected whole, i.e. a system rather than an aggregate. (Recall Definition 3.12 in Ch. 3, Sec. 2.7.) However, it is not a very tightly knit one.

The hypothesis that the world is a system is weaker, but presumably nearer the truth, than the organicist cosmology of Plato, the Stoics, or Hegel, according to whom every thing is bonded to everything else, so that the world as a whole is an "organic whole". If this other hypothesis were true then it would be impossible to study any particular thing without taking cognizance of every other thing, and it would be impossible to act on any particular thing without disturbing the entire universe.

Both contemporary science and our ontology steer a middle course between the extremes of organicism ("Everything hangs together forming a solid block") and atomism ("Everything goes its own way"). No part of the universe is isolated (or entirely free from bonds) but every thing is isolated in some respects from some other things. This *partial interconnectedness* of the parts of the world renders their study possible, for the study of every thing is partial and rests on the possibility of making contact with the thing.

Moreover, Postulate 5.10 supplies the existence criterion commonly used in the sciences, to wit: *Whatever exists (really, physically) is influenced by, or influences, some other things.* To repeat:

CRITERION 5.1 For any object x other than the world, *x exists really* iff there is a thing y other than x such that y acts on x or x acts on y.

If the thing y that acts on or is acted upon by the putative thing x, is controllable by an observer, then x is a subject of experimental research. But such possibility of effective control is sufficient, not necessary, to conjecture existence.

In short, our ontology includes the thesis of the limited or partial interconnectedness of all the parts of the universe.

5.4. *Three Misconceptions*

We have defined change with reference to concrete objects or things. Every collection of events is a set of changes in some thing or other. We have found no use for the fiction that there are changes that fail to consist in modifications of the states of some thing or other. Anyone claiming that there are such changes in reality should exhibit empirical evidence to this effect and proceed to build a theory of such thingless (or immaterial) changes.

Nevertheless once in a while philosophers and even scientists are found claiming that there are such thingless changes. Let us examine briefly the claims for three such items. One of them is the information process, sometimes regarded as not being based on the transport of matter or the propagation of fields. The reason for this misconception is that statistical information theory is a black box or phenomenological theory so general that it disregards the precise kind of signal carrying the information, the information transmission mechanism, and the kind(s) and quantity of energy involved. For this reason information theory is a nonphysical theory: it is instead a theory qualifying as a piece of scientific metaphysics, since it deals in an extremely general way with a genus of concrete things. But, of course, information is a property of certain physical (or chemical or biological) processes, namely signals, which are processes of a kind. No signal, no information transmission. And signals, let us repeat, are chains of events occurring in concrete things: the transmission of information consists in events propagating across space and carrying energy. Remove such real processes and only parapsychological anecdotes remain.

Another alleged example of a nonphysical yet real change is the parallelistic thesis that psychical or mental ("behavioral") events are not to be conceived of as changes in the nervous system of the animal concerned. If this is a point in methodology, it may have been sound in the beginnings of experimental psychology. But it has now become obnoxious for it blocks research in physiological psychology, and it is absurd when taken metaphysically, i.e. as restating the moth-eaten dogma that the soul is above matter. There is no evidence whatever for brainless ideation, whereas on the other hand there is definite evidence that every mental event is an event in the brain of some animal. We defer the detailed investigation of this matter to Vol. 4, Ch. 10.

A third point of incidence of immaterialism is the standard quantum mechanical theory of measurement (von Neumann, 1932). According to it the most striking difference between quantum physics and classical physics is that, while the latter treats measurement processes as purely physical processes, according to quantum mechanics they are not because of the intervention of the observer's mind and his decisions. So much so, it is alleged, that only natural processes obey the Schrödinger equation (but then they are unobservable) whereas measurement processes obey a separate theory centered on a different postulate, namely the projection postulate. This interpretation is not accepted by the orthodox followers of Bohr, according to whom *all* quantum events (which they call 'phenomena'), not just those controlled by measurement devices, involve observation acts. According to this alternative view there would be no spontaneous (unprovoked and unobserved) microevents ("quantum phenomena"). Realists reject of course Bohr's thesis as well as von Neumann's (Bunge, 1973b). One argument for the realist stand is that, while it is true that every measurement process is designed, executed and interpreted by some person (eventually with the help of automatic devices), it is a strictly physical process so far as the measured and the measuring things are concerned. So much so that (*a*) some measurement processes can be completely automated and (*b*) none of the existing measurement theories includes any mental variables. But even if the larger system including the observer and his staff had to be considered, this would not render measurement a nonphysical process – unless of course the mind were construed as an immaterial substance in the style of Aristotle and of the public (not the secret) Descartes. But modern physics need not ally itself with an obsolete metaphysics.

6. Concluding remarks

A basic polarity in traditional metaphysics is that of being and becoming: event is opposed to thing, process to stuff, change to structure. This opposition makes no sense in our system, where every change is the transformation of some thing or other, and every thing is in flux. This view, incompatible with much of traditional metaphysics, is consistent with science: the latter provides no ground for hypothesizing the existence of thingless events any more than it suggests that there might be changeless things. Surely the formulas of theoretical science do not always contain thing variables in an explicit way, but the context – what is sometimes called "the prose identifying the variables" – usually makes it clear that the formulas refer to things of some kind or other. The precise description of every event requires mentioning the thing or things subject to the change in question. This is why any correct axiomatization of a scientific theory starts by making assumptions concerning the kinds of thing the theory refers to, and ends up by characterizing events as changes in some of the properties of such referents. (See e.g. Bunge, 1967b, 1973b.) The converse course, namely that of starting with events and ending up by defining things, has been suggested as a program (e.g. by Russell (1914) and Whitehead (1929)) but never actually carried out. Let us take a glimpse at it.

Some process philosophers – anxious to flog the dead horse of Parmenidean ontology – have gone to the point of postulating that the being of an entity consists in its becoming – that "the actual world is a process" (Whitehead, 1929, pp. 29 and *passim*). Likewise the physicist Fokker characterized reality as a stream of events (Fokker, 1965, p. 1).

A few other physicists have reinvented this idea and even denied the existence of things (Bohm, 1970; Finkelstein, 1973; Stapp, 1976). This idea, that events and processes are the stuff the world is made of, has been encouraged by a misleading terminology introduced by some workers in relativistic physics, who call any point in spacetime an *event*, whether or not the point happens to be occupied by some thing. Further, they call *world* the set of all such "events" – even for a hollow universe. In this way the world is regarded as a set of events and the things that change are lost sight of.

In point of fact not the world but just a *world line* or tube or history – the region of spacetime scanned by some thing – qualifies as a sequence of events or process. (See Figure 5.19.) What can be said is that any region of spacetime is the seat of *possible* events occurring to things (particles, fields, etc.) that might exist within that region. But this is a far cry from identifying an event with a quadruple of real numbers.

In any case we need not spend more time with the quaint version of process metaphysics that dispenses with the thing concept, because it is logically untenable. Indeed, it is circular to define a *thing* as the

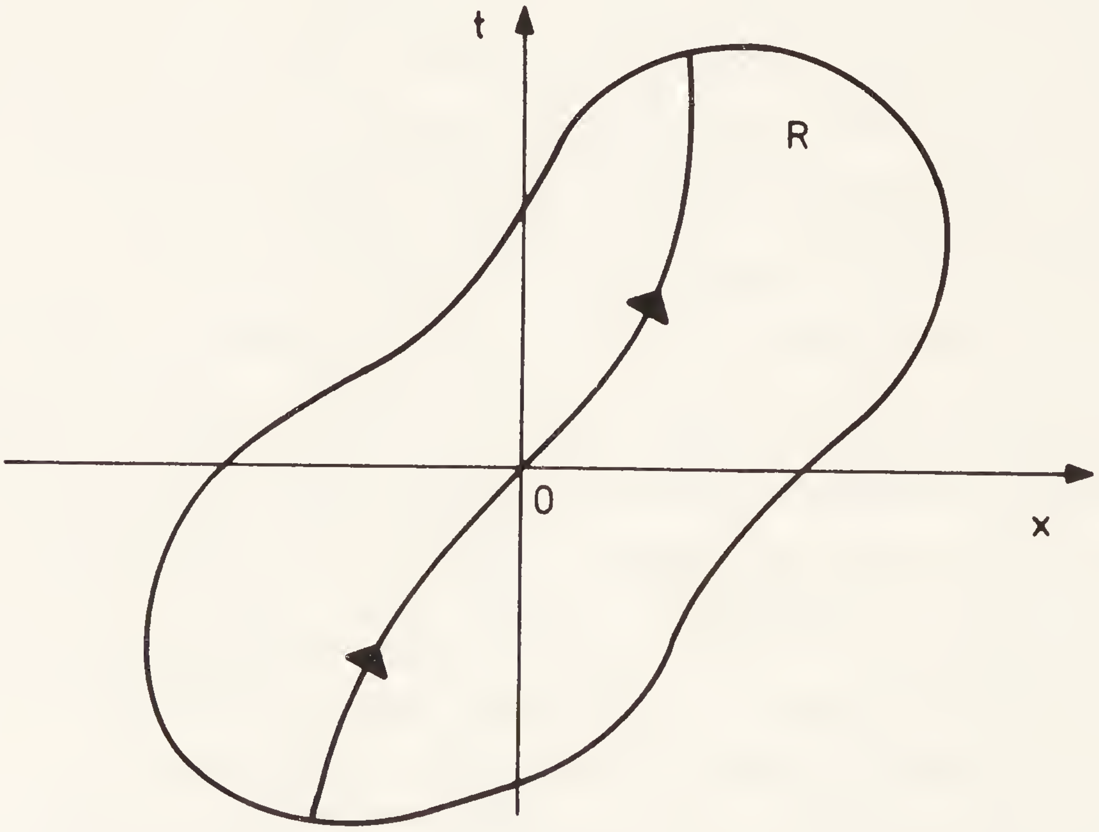


Fig. 5.19. The spatiotemporal track of a point particle (or world line of the latter) is a process in spacetime. Unless there are actually other things in the spatiotemporal region *R*, all the points other than those constituting the world line of the particle are uneventful and therefore do not deserve being called 'events'.

collection of events occurring in the *thing*. This is why our study of change was preceded by an examination of the concept of a thing.

The dual of process metaphysics is of course being metaphysics. One version of it is structuralism, currently fashionable in France (cf. Lévi-Strauss, 1958). Just as traditional metaphysics opposed being to becoming (or thing to process), and even extolled one of them at the expense of the other, so structuralism tends to oppose structure to change and to divorce the former from things. Just as the extreme dynamicists speak of thingless events, so structuralists speak of structures in themselves and concentrate their attention on structural permanence amidst flux. This too is mistaken: except for pure mathematics, where there are structures in themselves – e.g. Boolean algebras – every structure is the structure *of* some (complex) thing – atom, molecule, cell, organ, organism, community, ecosystem, or what have you. (Recall Definition 5.34 of the structure of an aggregation of things.) We do not say of a certain physical object that it *is* a rotation group but that a certain property of it (e.g. the energy) has rotation symmetry, or is invariant under rotations (members of the rotation group). Likewise we do not say of a community that it *is* a social structure but instead that it *has* such a structure. Moreover given the changeability of all things it is unlikely that there be anything with a permanent structure. In short, structure is a property – and a changeable one.

In conclusion, neither processism nor structuralism are viable alternatives to our metaphysics of changing things. Becoming is not pure flux but consists of things' being in successive states, or of successive things. (Remember Aristotle's *Physics*, Bk. III, Ch. 1, Sec. 1: "there is no such thing as motion *over and above* the things".) And structures are not just fickle but also lacking in independent existence: they are structures of things, features of concrete objects. In sum change and structure are so many distinct but interlocked features of changing things endowed with some structure or other. Therefore the traditional metaphysical poles, namely process metaphysics and being metaphysics, have each a grain of truth and both are overcome by our ontology.

We come to the end of our study of change in general. We have formed the main building blocks for constructing the concepts of time and space. Once these have become available we shall be able to give a more detailed account of change – in particular qualitative change, which will concern us in Vol. 4, Ch. 8. We proceed then to a study of extension and duration.

CHAPTER 6

SPACETIME

Science takes space and time for granted. Although not every scientific theory makes detailed assumptions about space and time, these two are usually regarded as constituting an independently existing framework allowing one to locate things and date events. So much so that facts are said to be or to happen *in* space and time even when we do not care for their precise spatiotemporal characteristics. So much are space and time taken for granted in science that most descriptions of things are effected in terms of space or time – i.e. space and time coordinates are used as independent variables. Moreover space and time are usually regarded as external to things and their changes: they are taken to constitute a fixed scenario.

True, on the relativistic theory of gravitation things influence the structure of spacetime: they distort its metric, hence its intrinsic curvature. But they do not create the basic manifold. (Indeed, setting the matter tensor equal to zero in the gravitational field equations one is left with equations describing a riemannian space.) In other words, even the general theory of relativity is consistent with – although it does not necessitate – a modified absolute theory of spacetime, i.e. one according to which spacetime has an independent existence instead of being a mesh of relations among facts. (The revision required by general relativity is this: the metric is not a priori, or purely geometric, but depends upon the momentary distribution of matter and fields. Hence spacetime is homogeneous just in case that distribution is itself homogeneous – i.e. actually never.) In sum science tells us what the structure of spacetime is but not what spacetime is.

This procedure, though legitimate in science, is unsatisfactory in philosophy. Here we cannot take space and time for granted because philosophy takes for granted nothing but the existence of the whole universe. In philosophy we must ask 'What are space and time?': we must attempt to understand how they come about, what their roots in things and events are. For, in the absence of things, there should be no spatial relations; and in the absence of change there should be no temporal relations. Indeed, it takes at least two things to make sense of 'here', 'there', 'to the left', and the like. And it takes at least two different states of a thing to make sense of 'before', 'after', 'meanwhile', and their kin.

Therefore instead of assuming that space and time are absolute (autonomous, self-existing) or even mildly absolute (influenced but not created by the furniture of the world), we should try and construct them out of facts. Philosophers are not interested in the empirical question 'Where and when did x happen?' They are interested instead in the question 'Do facts give rise to space and time, and if so how?' It behoves them to characterize the gross – algebraic and topological – structure of space and time, if possible in terms of factual items. Once this task has been accomplished, the philosopher should step aside and let the scientist investigate the fine structure – in particular the metric – of spacetime. It would be presumptuous of him to legislate that spacetime is – or is not – euclidean: this is a matter for the scientist to investigate. The philosopher will not try to compete with the scientist unless he is intent on being beaten by him.

From a methodological point of view the situation is this: there are three kinds of geometry – mathematical, physical, and ontological. *Mathematical geometry*, or geometry for short, is the collection of theories defining spaces of all kinds, where by 'space' is meant an arbitrary set endowed with a minimal structure, namely a topology. This structure can be modest or rich, and it may or may not depend upon the nature of the elements of the set. For example let S be a two member set $S=\{a, b\}$ – never mind what these elements are. Form the power set of S, i.e. $\tau=2^S=\{\varnothing, \{a\}, \{b\}, \{a, b\}\}$. Then S is said to be *topologized* by τ. Equivalently, the ordered couple $\langle S, \tau\rangle$ is called a *topological space* – or space for short. A theory of a mathematical space is said to be "concrete" if the nature of the elements of the base set S is specified – e.g. if the members of S are n-tuples of real numbers. Otherwise, i.e. if the "points" of the space are faceless (unspecified individuals), then the theory is said to be abstract, or to define an abstract space.

A *physical geometry*, on the other hand, is a theory purporting to represent things and events in physical space (or in spacetime). It consists of a mathematical geometry of the "concrete" type (i.e. interpreted in mathematical terms) enriched with semantic assumptions ("correspondence rules") specifying the referents of the theory (e.g. light rays) and what properties of such referents (e.g. their mutual separation) those concepts represent. For example, in such a geometry a (physical) triangle could be determined, say, by three intersecting light rays, and the length of its sides could be equalled to the times of flight of the photons constituting such rays. That is, the objects of a physical geometry are physical objects and as such subject to measurement at least in principle.

Finally we shall call *philosophical geometry*, or *chronotopics*, the set of theories explaining the deep structure, i.e. the factual support, of spacetime. (A philosophical geometry is a part of both philosophy – in particular ontology – and the general foundations of physics. It should not be mistaken for the philosophy of geometry, which is a metatheory concerned with the nature of geometric objects, the kinds of geometric argument, the relation of mathematical geometry to factual science and to reality, and so on.) Chronotopics makes use of both ontology and pure mathematics, and it intends to underpin physical geometry. It attempts to lay bare the ontic substratum of spatiotemporal relations in order to answer the philosophical questions that neither mathematical nor physical geometry answer, namely 'What are space and time?'. The subject of the present chapter is precisely chronotopics. And our answer to the question formulated a wink ago will be this: Space and time are what chronotopics say they are. Or, in a less sibylline way: *Spacetime is the collection of facts together with their separations* – where "separation" is construed in terms of facts only. Precisely how it is construed will be seen shortly. But before embarking on the construction of the theory we must examine the kinds of theory we might build.

1. CONFLICTING VIEWS

1.1. *The Three Main Views*

There are three main views on the nature of space and time: we call them the container, the prime stuff, and the relational views. (For a history of these ideas see Jammer, 1954.) Here they are in a nutshell:

(i) *The container view*: space and time constitute the fixed stage where things play out their comedy. Physical objects exist in space and time, which in turn are not physical objects. Moreover a physical object may be defined as anything occupying a region of space and a stretch of time. The container exists by itself (is absolute), hence it would continue to exist if all things contained were to change radically in kind or even were to pass out of existence. Not being a physical thing or a relation among physical objects, the supreme container cannot be described in physical terms: space and time have got to be described in purely mathematical terms without even the assistance of concept-fact bridges or semantic assumptions. Thus instead of stating that a certain geometry represents physical space we must say that the two are identical. Such a view is compatible with both the objectivist conception invented by the Greek atomists, and refined by Newton, and the subjectivist conception of space and time as the intuitive scaffolding necessary for human experience (Kant).

The container view of space and time is strongly suggested by the way science uses these concepts most of the time. Thus one calculates or measures the place or time *at* which something happens, and one computes or measures spatial and temporal intervals regardless of the kind of fact involved. However, the fixed framework view has been shaken by the relativistic theory of gravitation, where the very structure of spacetime is influenced (though not determined) by the things contained. (As you turn the pages of this book you modify, albeit by tiny amounts, the distances between neighboring objects as well as the trajectories of the light rays reaching your eyes.) Besides, the doctrine is philosophically unsatisfactory because it does not really answer – except metaphorically – the question 'What are space and time?'. It is particularly unacceptable to any ontology which, like ours, has no room for objects that are neither things nor properties of things nor relations among things. Therefore we cannot accept the container view even though we may feign to do so in everyday life.

(ii) *The prime stuff view*: spacetime is the elementary substance of which every physical object is made. Whatever is part of the world is a chunk of spacetime, and every substantial property is a property of a chunk of spacetime. Hence everything physical ought to be explained in spatiotemporal terms – be it Anaximander's *ápeiron*, Clifford's space hills, or Wheeler's wormholes. Physics becomes geometry and moreover a geometry in no need of physical interpretation, as it generates its own interpretation. Geometric monism is substituted for the spacetime-matter dualism of the container view: space and time are not just experientially and epistemically prior (as for Kant) but also ontologically primary (as for Alexander). Indeed, every physical object is regarded as a local warping of space (or of spacetime). Thus an electrically charged particle is said to be identical to a wormhole in the basic spacetime manifold. (See Wheeler, 1962.) In short, on this view space and time are just as absolute as on the container view, in so far as their properties depend on nothing. The difference is that, according to the prime stuff view, things are not in spacetime but instead the latter constitutes the former.

This fascinating theory is unfortunately semantically and ontologically too simple to be true. Semantically because the mathematical formalism is declared to be in no need of semantic hypotheses ("correspondence rules"). Thus instead of saying that a certain figure in spacetime *represents* a black hole one simply *calls* that surface a black hole: definitions take the place of semantic assumptions, hence the border line between formal science and factual science disappears – perhaps also that between constructs and things. And the theory is ontologically too simple because it attempts to reduce the entire variety of kinds of stuff to extension and shape (in four dimensions) alone. This Cartesian reduction to *figures et mouvements* does not account for the varieties of elementary particles and fields.

(iii) *The relational view*: space and time are not self-existing objects but a network of relations among factual items – things and their changes. What is assigned mathematical (topological, affine, or metrical) properties is not spacetime itself but the set of things – atoms, fields, etc. – and their changes. No changing things, no spacetime. Of course things may be said to have spatiotemporal properties – but the latter boil down to relations among things or events. The concepts of space and time used in science should result (and also be made more precise) upon one's abstracting from the facts, as when "Point x is included in sphere s with center at the origin of coordinate system k" is abstracted from "Thing X is inside ball S centered at the origin of the material frame of reference f (represented by coordinate system k)". Likewise "Instant t_1 precedes instant t_2 in the four-dimensional coordinate system k" should be construed as an abstraction from "Event e_1 precedes event e_2 in reference frame f (represented by the four-dimensional coordinate system k)".

The relational view has been expounded, with various degrees of clarity, by thinkers as diverse as Aristotle, Lucretius, Augustine, Leibniz, Lambert, Lobachevsky, Riemann, Engels, Mach, Whitehead, Robb, van Dantzig, Fokker, and Penrose. It was best summarized by Leibniz in oft quoted words: space is an *order of possible coexistents* and time an *order of successives* (Leibniz, 1956, II, p. 1083). Hence talk of space or time is elliptic talk of facts (Mach, 1872, 1883). When saying

'The position of thing X' we actually mean "The position of thing X relative to thing Y", where Y is either a reference frame or just the environment of X. And when saying 'The state of thing X at time t' we actually mean "The state of thing X when the chosen reference frame (or the environment) is in state t". However, *pace* Mach it does not follow that the concepts of space and time are dispensable: elucidating is not necessarily explaining away or eliminating. We cannot afford such a luxury because mention of space and time landmarks amounts to a public and global, hence objective and economical, reference to the state of the environment – of any environment, of any kind. Moreover there is no better way of getting quantitative precision than to fix the spatiotemporal coordinates of a thing or an event. But this is a methodological point: what matters to ontology is that space and time are not self-existing (absolute) objects of uncertain ontological status (neither things nor thing properties). The relational view is that *spacetime is the basic structure of the totality of possible facts*. But, unlike the prime stuff view, which eventually developed into a full fledged theory (geometrodynamics), the relational view has heretofore remained at the heuristic stage. The subsequent sections will see us through a hypothetical-deductive system spelling out the basic intuitions of the relationist thinkers. But before piling up bricks let us take a look at architectural styles.

1.2. *Approaches to Chronotopics Building*

There are two epistemological approaches to the problem of building a relational chronotopics: the subjectivist and the objectivist ones. In the former one takes the cognitive subject and his experiences as the point of departure, whereas in the second case one proceeds from a consideration of the things themselves. The subjectivist approach has been tried by Whitehead, Nicod and Basri among others. Let us glance at these attempts.

Whitehead (1919) assumes that space is euclidean and tries to define in experiential terms all the concepts that euclidean geometry takes as primitive. (There is an echo of the Berkeley–Mach program in Whitehead's attempting "the deduction of scientific concepts from the simplest elements of our perceptual knowledge": Whitehead, 1919, p. vii.) Thus Whitehead's is not really a radical construction but a reconstruction with a strong a priori component, namely a particular mathematical conponent, and an equally strong philosophical bias, namely radical empiricism. Because of that a priori element Whitehead's theory of space and time is incompatible (by design) with the two theories of relativity and therefore hopelessly obsolete at birth and beyond repair. And because of its pronounced subjectivism the theory is inconsistent with the epistemological outlook of science, which is realist and the one adopted in this work. (Recall our Rule 7 for philosophizing: Introduction, Sec. 4.) Lucas' more recent attempt (Lucas, 1973) is subject to similar criticisms.

Nicod's (1923) venture was more radical though technically far less accomplished than Whitehead's and just as far removed from the science of his time. It was more radical because Nicod started from (imaginary) experiences and attempted to build geometry out of them without imposing any a priori structure on space. (One is reminded of Condillac and his statue.) In other words Nicod – following leads from Mach and Russell, and independently of Carnap's similar attempt – tried to build physical geometry out of (common sense) psychology. As a matter of fact he delivered nothing of the sort: he did not even produce a theory of any of the perceptual spaces such as the visual or the acoustic space. (Neither he nor Whitehead realized the differences between physical space and the perceptual spaces.) Like Whitehead's, Nicod's was a scientific and philosophic failure. (For detailed criticisms see McGilvary, 1951 and Vuillemin, 1971.)

Finally the most recent work along the Mach-early Russell–Whitehead–Carnap line, namely Basri's (1966), is just as subjectivistic but on the other hand it has the merit of being admirably rigorous and of respecting the formulas of relativistic physics. However, it does not serve as a foundation for the latter because it violates the spirit of the two relativities, which are field theories (Basri admits only particles) and moreover observer-invariant ones (Basri's is observer-bound). And although this theory professes to be concerned with human observers and their sensations and operations, it makes no contribution to the physiology or the psychology of space and time perception. In short, since the theory is subservient to an unscientific philosophy – namely empiricism – it serves neither physics nor psychology.

The subjectivistic approach cannot yield what we want in scientific ontology, namely an objective chronotopics compatible with physics but soft enough that the physicist may shape it according to his theories and data. And, whether subjectivist or merely conventionalist, the a priori approach to the geometry of physical space is unphysical because it assigns space a structure independent of the actual distribution of things. We must therefore try an objectivist approach, i.e. one starting with factual items as the prime stuff and postulating their spatiotemporal relations quite apart from considerations of perception and even measurement. This does not mean that the construction should proceed a priori, i.e. in advance of any experience. Quite the contrary, we shall be guided by the finding of scientific experience, that physical space is a three-dimensional connected generalized continuum. For the sake of simplicity we shall build first a geometry, then a chronology, and finally a chronotopics or philosophical theory of spacetime.

2. Space

2.1. *Interposition*

We shall build our relational theory of space using only materials elaborated in the previous chapters. In fact we need only the following background assumptions and definitions:

(i) there are concrete objects or things, none of them is a construct, and all of them exist independently of the cognitive subject;

(ii) all things can juxtapose or aggregate: if $x, y \in \Theta$, then $x \dotplus y \in \Theta$ or, equivalently, $[\{x, y\}] \in \Theta$;

(iii) the world is the aggregation of all things – i.e. $\square = [\Theta]$ – and whatever is a part of the world is a thing – i.e. if $x \sqsubset \square$ then $x \in \Theta$;

(iv) all things are, or are composed of, basic things or members of $B \subset \Theta$: for any $x \in \Theta$ there is a unique subset B_x of B such that $x = [B_x]$; and basics are elementary, i.e. have no parts: if $x, y \subset B$ and $x \sqsubset y$, then $x = y$.

(v) every thing is in some state or other relative to a given standard thing or reference frame, and the collection of really possible (lawful) states of a thing x is called its lawful state space $S_{\mathbb{L}}(x)$;

(vi) the evolution or history of every thing x is representable as a trajectory in its lawful state space: if $\mathbb{F}$ is the state function of x, and $S(f)$ the state space of reference frame f, then $h(x) = \{\langle t, \mathbb{F}(t)\rangle | t \in S(f)\}$;

(vii) a thing x acts upon another thing y iff the former modifies the history of the latter: $x \rhd y =_{df} h(x \dotplus y) = h(x) \cup h(y|x)$ & $h(y|x) \neq h(y)$;

(viii) the total effect or action of thing x upon thing y is $A(x, y) = h(y|x) \cap \overline{h(y)}$;

(ix) every change is a change in the state of some thing, and the net change of thing x from state s to state s' is representable as the ordered pair $e = \langle s, s' \rangle \in S_{\mathbb{L}}(x) \times S_{\mathbb{L}}(x)$;

(x) if $e = \langle s, s' \rangle$ and $e' = \langle s'', s''' \rangle$ are events in a thing x relative to a frame f, then e precedes e' if e composes with e' in the indicated order to form a third event in x; i.e. $e < e' =_{df} e * e' \in E_{\mathbb{L}}(x)$, where $E_{\mathbb{L}}(x)$ is the proper lawful event space of x, i.e. the set of really possible changes of x.

Our first geometrical notion is that of interposition, which we proceed to build in terms of the ontological notions occurring in the previous list. Our definition will be explicit, namely via

POSTULATE 6.1 Let R be a ternary relation among things and abbreviate "R holds among x, y, z in the given order" to $x|y|z$, where $x, y, z \in \Theta$. Then R is the *interposition* or *betweenness* relation iff, for any $u, v, x, y, z \in \Theta$,

(i) $x|y|z \Rightarrow x \neq y \neq z \neq x$ or $x = y = z$, i.e. R holds either among different things or trivially for a single thing;

(ii) $x|y|z \Rightarrow z|y|x$, i.e. R is symmetric in the outer variables;

(iii) $x|y|z \;\&\; u|x|y \;\&\; y|z|v \Rightarrow u|y|v$, i.e. whatever interposes between two given things lies also between two outer things;

(iv) $x \neq y \;\&\; x \sqsubset y \Rightarrow \neg(\exists z)(z \in \Theta \;\&\; x|z|y)$, i.e. nothing interposes between the part and the whole;

(v) $y \in B \Rightarrow (\exists x)(\exists z)(x, z \in B - \{y\} \;\&\; x|y|z)$, i.e. any basic thing can be "surrounded" by two other basics;

(vi) $\neg(x \sqsubset y) \;\&\; \neg(y \sqsubset x) \Rightarrow (\exists z)(z \in B \;\&\; x|z|y)$, i.e., there are basics interposed between any two detached things: the world is dense, it constitutes a *plenum*;

(vii) $x|y|z \;\&\; x \rhd z \Rightarrow (\exists u)[u \in A(x, y) \;\&\; (v)(v \in A(x, z) \Rightarrow u < v)]$, i.e. if y interposes between x and z, and x acts on z, then some of the effects of x on y precede all of the effects of x upon z. (See Figure 6.1.)

The first three clauses seem intuitive. The fourth is a sort of interpretation of Hilbert's second axiom group for elementary geometry – in our notation $\neg(x|y|y)$ (Hilbert, 1899). The fifth is a sort of dual of Hilbert's 4th component of the second axiom group – namely that there is a thing lying between any two distinct things. The sixth is countenanced by both gravitation theory and quantum electrodynamics: according to these theories no region of space is totally devoid of some entity or other. (This is of course the plenistic hypothesis defended by Aristotle, Descartes and Einstein among others.) And the last clause exhibits,

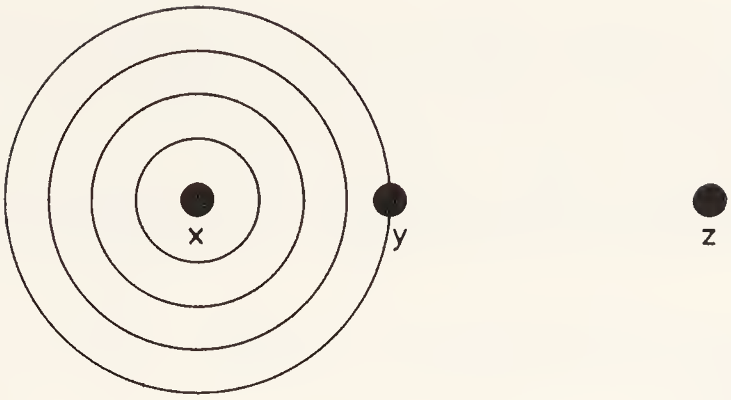


Fig. 6.1. Illustration of clause (vii) of Postulate 6.1. A spherical wave emitted by x reaches y before it does z.

perhaps even more forcefully than all the others, the material basis of the interposition relation. It would be objected to by the Leibnizians, who would argue that there would be space even in a world composed of mutually independent entities (monads), none of which acted upon any other. Maybe, but ours happens not to be such a world (Postulate 5.10). Clause (vii) suggests that the interposition relation might not even be properly definable for a changeless universe.

2.2. *A Philosopher's Space*

A basic concept in any geometry is that of separation. There are several notions of separation. One of them is the topological notion of separation between sets (Wallace 1941). We cannot use it because we wish to clarify the notion of separation between concrete things not between sets, which are constructs. Other concepts of separation are metrical or quasimetrical. Any of these consists in a real valued function d on an abstract set S obeying well known conditions. The members x and y of S are said to be separate iff $d(x, y) \neq 0$. We cannot use these notions either because it is not our purpose to compete with physicists by assigning precise quantitative measures to the separations among things. We need a more basic, qualitative notion of separation. We shall get it with the help of the concept of interposition.

We define the separation between two things as the set of things that interpose or lie between the given things. More precisely, we make

DEFINITION 6.1 Let $B \subset \Theta$ be the set of basic things. The function $\sigma: \Theta \times \Theta \rightarrow 2^B$ such that $\sigma(x, y) = \{z \in B \mid x|z|y\}$ for $x, y \in \Theta$ is called the *thing separation*.

Because the null thing is part of every thing, the separation between an arbitrary thing and the null thing is nought, i.e. $\sigma(x, \square) = \varnothing$. And because every thing is part of the world, there is no separation either between an arbitrary thing and the world: $\sigma(x, \text{▯}) = \varnothing$. These two consequences follow from Postulate 6.1 and Definition 6.1, as does

THEOREM 6.1 For any two things $x, y \in \Theta$,
(i) $\sigma(x, x) = \{x\}$;
(ii) $\sigma(x, y) = \sigma(y, x)$.

Proof The first part follows from clause (i) of Postulate 6.1. The second, from clause (ii).

Note the similarities and dissimilarities between σ and the distance (or quasidistance) functions d mentioned at the beginning of this section. Property (ii) is analogous to the corresponding properties of the d's. On the other hand property (i) is not, nor is the frame invariance of σ, which contrasts with the frame dependence of the d's.

Further, if a thing interposes between two things which in turn interpose between two other things, then the former interposes between the latter. In terms of the separation function:

THEOREM 6.2 For any things $u, v, x, y, z \in \Theta$: If $y \in \sigma(x, z)$, $x \in \sigma(u, y)$ and $z \in \sigma(y, v)$, then $y \in \sigma(u, v)$.

Proof By Postulate 6.1(iii) and Definition 6.1.

We could derive further theorems but we shall not need them for our purposes. What we do need is another property of the separation function, which the previous assumptions and definitions fail to entail. We must therefore postulate it (see Figure 6.2). More exactly, we assume

POSTULATE 6.2 Let x_1, x_2, y, z_1 and z_2 be basic things. If $y \in \sigma(x_1, z_1) \cap \sigma(x_2, z_2)$ and $x_1 \neq z_1$, $x_2 \neq z_2$, then there are basic things x_3 and z_3 such that $x_3 \neq z_3$ and

$$y \in \sigma(x_3, z_3) \subset \sigma(x_1, z_1) \cap \sigma(x_2, z_2).$$

We now have all we need to build an important notion:

DEFINITION 6.2 The set B of basic things, together with the separation function σ, is called the thing space, abbreviated $\vartheta = \langle B, \sigma \rangle$.

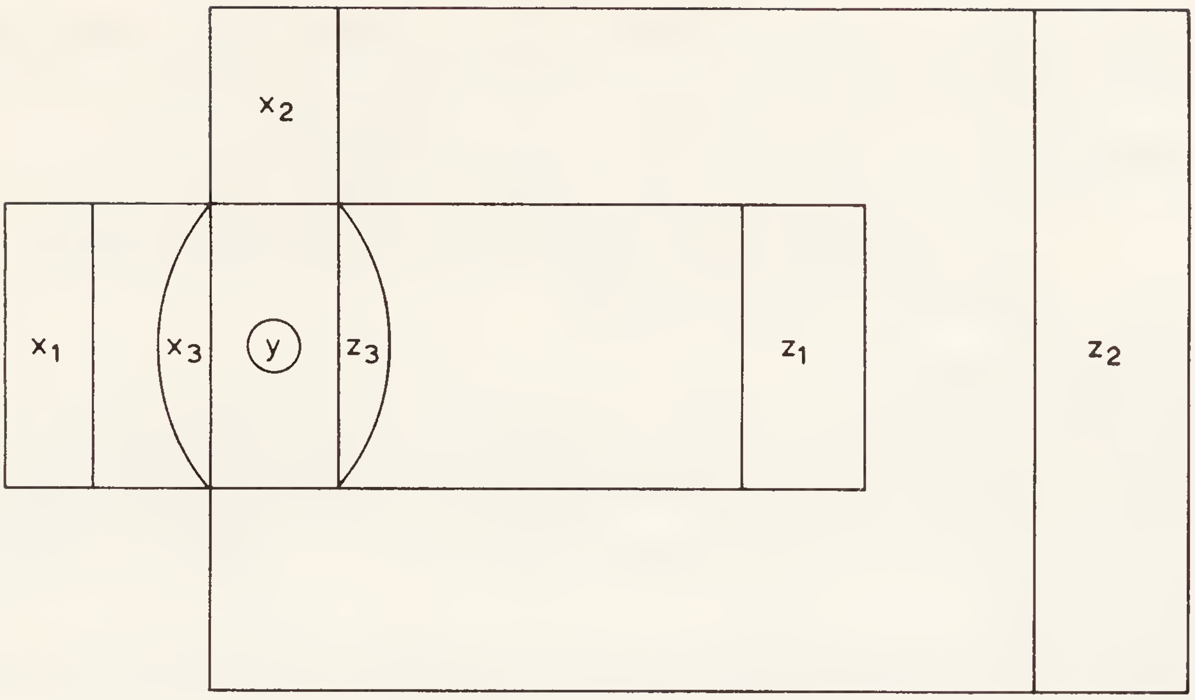


Fig. 6.2. Postulate 6.2.

In other words, the thing space is nothing but *the collection of spaced things*, or the set of things related by their mutual separations. This space deserves being regarded as a real (or ordinary or physical) space because its coordinates, B and σ, are real. Indeed, the basics have been assumed to be real and moreover the fundamental constituents of all concrete objects; and the separation function is a *relatio realis* in the sense of Ch. 2, Sec. 5.1, Definition 2.17, for σ obtains between real things.

The space ϑ is all the philosopher needs, for it has been built out of a few extremely general (ontological) concepts. We may call ϑ a *philosopher's space*, and the corresponding theory (summed up in the preceding postulates, definitions and theorems), a *philosophical geometry*, and moreover one both relational and objectivist. However, so far we have no evidence that ϑ is indeed the physical space: we have not proved that our philosophical geometry can be specified to yield a geometry utilizable by physics. This task will be taken up in the next subsection, adapted from Bunge and García Máynez (1977).

2.3. *The Physicist's Space*

The thing space introduced by Definition 6.2 would seem to have too poor a structure to qualify as physical space or even as a mathematical space. However, this is not so, as will be shown in the present subsection.

To begin with $\langle B, \sigma\rangle$ has a definite topological structure, as shown by

THEOREM 6.3 The family of sets of basic things

$$\tau = \{X \in 2^B \mid \text{For all } y \in X \text{ there are } x \text{ and } z \text{ in } B - \{y\} \text{ such that } y \in \sigma(x, z) \subset X\}$$

is a topology for B.

Proof By Postulate 6.1(v), $B \in \tau$. Besides, it is obvious that $\varnothing \in \tau$ and that every union of members of τ belongs to τ. Finally, by Postulate 6.2 the intersection of any two elements of τ is also in τ. Hence τ is indeed a topology for B.

We shall assume then that the set B of basic things, equipped with the topology τ, is or rather represents physical space:

POSTULATE 6.3 The topological space $\Sigma = \langle B, \tau\rangle$ represents physical (ordinary, real) space.

However, we have so far little justification for assuming this hypothesis aside from the fact that B is constituted by things and has a topology. For one thing we have assigned Σ no definite dimensionality, whereas we feel certain that physical space is three-dimensional. For another we want Σ to be connected (in one piece). These and other shortcomings will be remedied by foisting certain properties on Σ. That is, we shall *force* Postulate 6.3 to become true.

To begin with we define, in the usual way, closures in τ:

DEFINITION 6.3 If $A \in \tau$, then

$$Cl\,A =_{df} \{x \in B \mid \text{For all } V \text{ in } \tau \text{ containing } x,\ V \cap A \neq \varnothing\}.$$

For these closures to be of any help they must be large enough. This is ensured by a mild enough assumption:

POSTULATE 6.4 Every pair of separate basics is in the τ-closure of their separation. I.e.,

$$\text{For any } x, y \in B, \text{ if } \sigma(x, y) \neq \varnothing, \text{ then } x, y \in Cl\,\sigma(x, y).$$

By taking $x = y$ it follows immediately that every finite subset of B is closed. Now, a space with this property is called a T_1-space. In other words, we have

COROLLARY 6.1 The physical space $\Sigma = \langle B, \tau \rangle$ is a T_1-space.

We need one more assumption to convert Σ into a Hausdorff (or T_2) space. Actually Σ should have further properties: the physicist wants Σ to be a connected and metrizable space locally homeomorphic to $\mathbb{R}^3$. In other words Σ should be a Euclidean three-manifold.

Now, it has been known for a long time that it is possible to express every theorem in Euclidean three-dimensional geometry with the sole help of the concepts of sphere and of inclusion. In fact these are the sole extralogical undefined notions in the set of postulates for elementary geometry proposed by Huntington (1913). For example, a *point* is defined in this geometry as a sphere such that there is no other sphere within the given sphere. And a *segment* is defined as follows: Let a and b be any given points, i.e. minimal spheres. If x is a point such that every sphere which contains both a and b also contains x, then x is said to belong to the *segment* (ab) or (ba). (See Figure 6.3.)

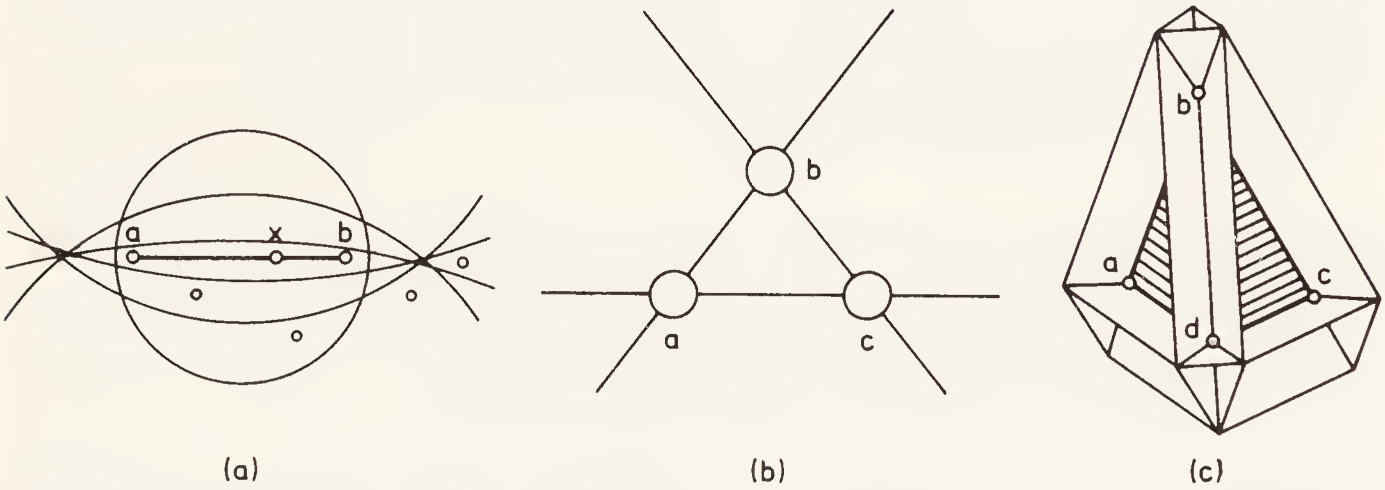


Fig. 6.3. Building space with spheres: (a) a segment, (b) a triangle, (c) a tetrahedron. From Huntington (1913), figs. 1, 3, 9.

We shall adopt Huntington's postulates but, instead of taking the notion of sphere as primitive, we shall define it in terms of some of our previous concepts. This will endow Huntington's postulates with a physical (or rather protophysical or ontological) interpretation and authorize us therefore to claim that they characterize physical space (in the small).

Our first task is then to define the concept of a sphere. We will define a sphere lying between two given things as the closure of a separation between those things. More precisely we make

DEFINITION 6.4 For any pair $x, z \in B$ of distinct basic things, the *spheres* lying between x and z are

$$Sxz = \{Cl\,\sigma(u, v) | u, v \in B \ \& \ \varnothing \neq Cl\,\sigma(u, v) \subset \sigma(x, z)\}.$$

Next we assume that these spheres satisfy Huntington's postulates:

POSTULATE 6.5 For any given pair $x, z \in B$ of distinct basics, if the spheres Sxz are nonempty, the structure $\langle Sxz, \subset \rangle$ satisfies (is a model of) the Huntington (1913) axioms.

Hence

THEOREM 6.4 The nonempty separation $\sigma(x, z)$ between two basic things $x, z \in B$ is homeomorphic to three-dimensional Euclidean space.

Proof The minimal spheres are the single points of $\sigma(x, z)$. By Huntington's (1913) Theorem 47 this collection can be topologized in such a way that it is homeomorphic to $\mathbb{R}^3$. Now, a basis for this topology consists of all the τ-interiors of elements of Sxz. Hence this latter topology is identical with the relative topology of $\sigma(x, z)$.

We are now justified in framing

DEFINITION 6.5 Every member $Cl\,\sigma(u, v)$ of the family of spheres Sxz lying between the things x and z is said to be a *Euclidean ball*.

Remember now that physical space is a T_1-space (Corollary 6.1). Since Postulate 6.5 supplies regularity we obtain

COROLLARY 6.2 $\Sigma = \langle B, \tau \rangle$ is a regular three-manifold without boundary.

We have come near the goal. We shall attain it by adding one last assumption, namely that any two basic things can be bridged by a string of spheres (Figure 6.4). More precisely, we assume

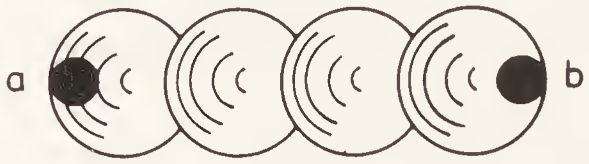


Fig. 6.4. The chain of partially overlapping spheres bridging things *a* and *b* (black dots).

POSTULATE 6.6 The set B of basics contains two sequences, $x_1, x_2, \ldots$ and $z_1, z_2, \ldots$, such that $x_i \neq z_i$ for each $i \in \mathbb{N}$ and, for every pair a, b in B, there is a simple chain

$$\langle C_i | 1 \leqslant i \leqslant n \rangle, [\text{i.e. } C_i \cap C_{i+1} \neq \varnothing, 1 \leqslant i \leqslant n-1]$$

such that $a \in C_1$ and $b \in C_n$, and every C_i is of the form $\sigma(x_j, z_j)$.

We can now prove that Σ is connected, i.e. in one piece or gapless. Moreover we shall prove in the same breath that Σ is second countable, i.e. that it has a countable basis. (A basis for τ is a subset β of τ such that each member of τ is a union of members of β, so that the latter are the building blocks of τ.) In fact we can also prove that Σ is metrizable. All of this is stated by

THEOREM 6.5 $\Sigma = \langle B, \tau \rangle$ is a connected, second countable and metrizable three-manifold without boundary.

Proof Being homeomorphic to $\mathbb{R}^3$ (by Theorem 6.4), every $\sigma(x_i, z_i)$ is connected and Lindelöf. Hence by Postulate 6.6 $\langle B, \tau \rangle$ is connected and Lindelöf. Now, every regular Lindelöf space is paracompact and every paracompact locally metrizable space is metrizable. Therefore $\langle B, \tau \rangle$ is connected, Lindelöf and metrizable. Finally $\langle B, \tau \rangle$ is second countable because, in metrizable spaces, the properties of being Lindelöf and of being second countable are equivalent. [For these notions see e.g. Gaal (1964).]

In other words, physical space is a three dimensional connected manifold. And, because Σ is metrizable, the physicist may assign it an adequate metric. We shall of course abstain from doing so because our goal was to build a concept of physical space broad enough that it can be used in the axiomatic foundations of any of the current physical theories. (Besides, the theories of relativity have taught us that any attempt to assign space a metric, independently of time, is bound to fail: only spacetime can be assigned the proper metric.)

This completes our justification of Postulate 6.3, that $\Sigma = \langle B, \tau \rangle$ represents physical space. But why should we have required Σ to have precisely the properties postulated or deduced in the present subsection? Why could we not hypothesize an entirely different set of properties of physical space? The answer to these questions depends upon the relation of one's philosophy with science, hence indirectly with reality.

In an a priori philosophy, physical space can be assigned any structure whatsoever. For example one could postulate – as Whitehead (1919) and Lucas (1973) have done – that physical space is globally euclidean. But in a philosophy that seeks to be continuous with science no such freedom exists. In any science-oriented philosophy one requires the geometry of ordinary space to agree with physics. Moreover one lets physics have the upper hand because physicists, not philosophers, are competent to decide on matters of the fine structure of physical space.

Now, physics happens to assume that physical space is a three-dimensional differentiable manifold. (See e.g. Trautman, 1965.) This is necessary to write down the basic equations of contemporary physics, though usually insufficient to solve them. (We disregard the speculative deviant theories. If any of them were shown to be true then we would have to change our geometry.) Surely most theories require some additional structure. For example classical mechanics and non-relativistic quantum mechanics require ordinary space to be globally euclidean, classical electrodynamics requires it to be an affine space, and the relativistic theory of gravitation assumes that space is riemannian. But the assumption common to all of these specifications is that physical space is a three-dimensional manifold. This being the minimal geometrical assumption, it should be a necessary and sufficient constraint for philosophy – until further notice. (Recall Rule 10 of the method of scientific ontology: Introduction, Sec. 4.)

The geometry described in the present subsection is then compatible with the mainstream of contemporary physics. A new physics might call for a new protophysical (or ontological) geometry. Two radical changes of the sort have been suggested a few times. One is that the spatial generalized continuum may have to be replaced by a discontinuous or atomic space with a fundamental length built into it. (This hypothesis is suspect because it usually has an operationalist motivation, namely the impossibility of measuring distances below a certain value, say 10^{-11} cm.) Another possibility is that our current metrics may be sorts of statistical averages of stochastic metrics describing a fluctuating spacetime. (This other hypothesis is far more plausible in view of the zero-point fluctuations of the electromagnetic field.) However none of these ideas seem to have been carried beyond the hand-waving stage. The fact is that the theories actually employed by physicists make no assumptions about physical space contradicting our results.

In sum, we may declare our ontological geometry true because contemporary physics says so. But at the same time we should be prepared to see it revised or even revolutionized by new developments in physics.

2.4. *Bulk and Shape*

Now that we have grounds for modelling physical space as a manifold we can use the latter to determine the bulk and shape of things. First we exactify the notion of the *place occupied by a thing*:

DEFINITION 6.6 Let the three-manifold $M^3 = \Sigma = \langle B, \tau \rangle$ represent physical space. Then the function $\beta\colon \Theta \to 2^{M^3}$ from things to regions of space is called the *bulk*, and its value $\beta(x)$ at $x \in \Theta$ the *bulk of* x, iff

(i) β is injective;
(ii) for every thing x other than the null thing, $\varnothing \subset \beta(x) \subseteq M^3$;
(iii) for all $x, y \in \Theta$, $\beta(x \dotplus y) = \beta(x) \cup \beta(y)$.

The first clause makes room for things, such as photons, that can share the same region of space. The second states that every real thing has a nonempty bulk – even if, as in the case of the mythical point particle, the bulk consists of a single point of space. The third, that β is additive. The three clauses together elucidate the notion of *room* or *space occupied by a thing*.

None of the above conditions presupposes or implies that all things have sharp boundaries: so far as Definition 6.6 is concerned, a thing can have "colonies", i.e. pieces outside its main bulk. Consequently nothing is implied about the measure of the bulk of a thing. Questions of measure are specific scientific questions answered by computing or measuring volumes not bulks. (And volumes, unlike bulks, depend on the interactions among the components, as well as on the reference frame.) On the other hand Definition 6.6 allows one to compare bulks. Indeed if x and y are things then we can say that x is *bulkier* than y just in case $\beta(x) \supseteq \beta(y)$.

The concept of bulk allows us to define that of shape:

DEFINITION 6.7 Let $x \in \Theta$ be a thing with bulk $\beta(x) \subset M^3$. Then x has *shape* f iff the boundary $\partial\beta(x)$ of $\beta(x)$ is homeomorphic to a piece-wise smooth function $f\colon \mathbb{R}^2 \to \mathbb{R}$.

Remark 1 The values of the shape function f depend upon the choice of reference frame, hence of coordinate system. Hence shape is frame-dependent rather than frame-invariant. (Recall the flattening of a sphere into an ellipsoid relative to a moving frame.) *Remark 2* Shape is not a universal property of things. For example electrons have no shape of their own – or, if preferred, they acquire the shape of whatever macrophysical body contains them. Nor do porous bodies and gaseous bodies have a definite shape of their own. Therefore we cannot regard shape as a basic property of all things. Hence we cannot adopt Descartes' program of explaining the world in terms of *figures et mouvements* only. *Remark 3* Shape, hardly a property of basic things, emerges rather definitely at the macromolecular level and becomes the more definite, the bulkier the thing. It is therefore a derivative property. Moreover it emerges from nongeometric characteristics. Thus the helicoidal configuration of a DNA molecule results from chemical forces such as the hydrogen bonds between an NH group and a carbonyl group, and it is influenced by the environment of the molecule – to the point that the pattern disintegrates at high temperatures. Likewise the shape of a macrobody is determined jointly by the inner stresses and the external forces. In general, shape or geometric pattern, when it exists at all, is an outcome of the interplay of internal forces and environmental constraints. *Remark 4* Although shape is a secondary property, once acquired it conditions the acquisition or loss of further properties, which are called *steric properties*. Suffice it to recall that the specific activity of enzymes depends upon their shape.

2.5. *Concluding Remarks*

We have obtained a solution to the problem we set out to solve, namely that of building a theory of physical space based on the concept of a changing thing. This does not mean that changing things are more fundamental than space: only the corresponding categories are ordered. Ours is then a full fledged relational theory – not just a program – and moreover one based on general ontological ideas and couched in strictly objective terms, as enjoined by Rule 7 of our philosophical methodology. (See Introduction, Sec. 4.)

However, ours is not the only possible relational theory of space. For example there is Basri's (1966) theory. But we rejected it in Sec. 1 for being a particle theory and therefore inconsistent with field physics, and also for being subjectivist, hence inconsistent with the epistemology inherent in the scientific approach. There is also Penrose's (1971) equally rigorous and moreover objectivist relational theory of space. But, because it is based on the concept of angular momentum, it is not general enough to serve as a foundation for all physical theories. On the other hand our own theory is based on extremely general ideas.

Our theory has one seemingly unsatisfactory feature, namely that it postulates the three-dimensionality of space instead of explaining it. (The hypothesis is contained in Postulate 6.5.) This assumption may seem arbitrary. Actually it is not, for it is suggested by the actual behavior of real things: if things were different then physical space might not be three-dimensional. In other words, the three-dimensionality of space is rooted in the laws of things. What would the dimensionality of physical space be if things possessed ("obeyed") laws different from those we know? This question can receive a partial answer, namely along the following lines. (For a detailed investigation see Penney (1965).) Take for instance the pervasive processes of propagation of waves of various kinds and write out any wave equations in spherical coordinates in an n-dimensional manifold. Any such equation will contain a radial term with a factor $(n-1)$. So, if experiment were to show that wave propagations proceed according to say, $n=4$, then we would have to conclude that physical space is four-dimensional. But of course this is not the case. In short experiment points, in however devious a fashion, to the three-dimensionality of ordinary space. (Note by the way that, on a nonrelational theory of space, one would ask instead the question: 'What would things look like if space were not three-dimensional?' But this question cannot be answered except dogmatically.) Hence in postulating that space is three-dimensional our ontology bows to experience, just as Peirce wanted.

What would become of space if all things were to go out of being? The answer is provided by the separation function (Definition 6.1). Since σ is defined only for pairs of things, and its values are sets of things, unless there are (changing) things there is no physical space, hence neither a relational philosophical geometry nor a physical geometry. That is, a hollow universe would be spaceless. Likewise a block universe, such as Parmenides' undifferentiated One, would be spaceless.

This corollary of our philosophical geometry has some bearing on the interpretation of the relativistic theory of gravitation. What happens if the matter tensor in the gravitational field equations is set equal to zero everywhere? Sure enough the solutions characterize a riemannian manifold: there would seem to be space even in the absence of things (particles, fields, etc.). Is this a physically possible solution? Not on our geometry: here, if there are no things, none of the relations among points in the manifold is a real relation. The relations involved in a riemannian geometry (e.g. the distance relation) become real to the extent that they are satisfiable by things. In short, the homogeneous equations of gravitation, corresponding to a hollow universe, are physically meaningless. (For the notion of factual content see Vol. 1, Ch. 5, Sec. 3 of this *Treatise*.) It is nice to see that ontology can be of some use to science.

Ordinary space, then, is just as real as any other real relation. (Actually space is not a relation but a set of related things or, if preferred, a relational structure: recall Definition 6.2.) But, not being a thing, physical space has no causal efficacy. In other words, the spatial relations are nonbonding relations rather than bonds or couplings. (For the concepts of bonding and nonbonding relations see Ch. 5, Sec. 4.1.) That is, just as things do not act upon space (since space is not a thing), so space does not react back on things. This is one point our theory shares with Newton's absolute view.

To the practising physicist, who takes space for granted, space is basic and things are somehow embedded in or superimposed on space. To the philosopher upholding a relational doctrine of space, neither is the more basic. In particular, on our theory things come with their mutual separations, hence neither is prior to the other: there is neither a thingless space nor a spaceless thing. This is of course in sharp contrast to the container view of space, according to which space, though not a thing, exists by itself and independently of the things it contains. According to our theory neither space nor things exist by themselves. Only mutually spaced things exist.

Moreover, the separations or spacings among things may alter with changes in the things themselves. Hence real space is as much in flux as are things. Real space is then a dynamic structure of the collection of things. Ordinary space may be pictured as an elastic net, or fluctuating lattice, the nodes of which are things. But temporality belongs in the next section.

3. Duration

3.1. *Intuitive Idea*

A duration is the duration of some event or process: a changeless universe would be timeless. Just as space is the spacing of things (Sec. 2), so time is the pace of events. (Aristotle, *Physics* IV, 11, 220a 25, called time *arithmós kinéseos*, the measure of motion.) And just as spatial distance is the separation among things, so temporal interval is the separation between different states. This is the intuitive germ of the relational theory of time.

To get an idea of this theory consider a world consisting of a single thing the states of which can be mapped on the set of natural numbers: it can be a clock or a heart. The history of such a heart is, roughly, the sequence of its heart-beats; and the life span of the heart is the total number of its heart-beats. Since there are not other hearts (or clocks) in our imaginary universe, there is no question of slowing down or quickening the pace of the heart-beats – not even of missing a heart-beat now and then. The heart-beats generate a uniform or well tempered time – i.e. same intervals between beats. The time coordinate attached to this clock is a function $t: B \to \mathbb{N}$, where B is the set of heart-beats and $\mathbb{N}$ that of natural numbers, such that $t(b) = n \in \mathbb{N}$ for $b \in B$. We interpret '$t(b)$' as the time at which b occurs. This function determines the time lapse function

$$\tau: B \times B \to \mathbb{N} \quad \text{such that } \tau(b, b') = |t(b) - t(b')| = |n - n'|$$

for $b, b' \in B$. (Clearly, the structure $\langle B, \tau \rangle$ is a metric space.) If the heart stops beating altogether then both t and τ lose their support and time ceases to "exist", i.e. the temporal relations cease to be real (to hold among real events). If we now multiply the number of time keepers it becomes possible to compare their paces and eventually to pick the most regular among them. But this is a technical point of interest to physics and astronomy rather than ontology. So much for the intuitive background.

3.2. *Before and After*

On a relational theory of time we cannot accept a purely mathematical characterization of time – e.g. as the ultimate independent variable, or as the parameter of a one-parameter continuous group of transformations. In other words there is no such thing as mathematical time in our ontology: there are only mathematical representations of time. (See however Whitrow (1961) for the concept of mathematical time.) Mathematics, unlike mathematicians, is alien to time: constructs are neither changeable nor mutable (Corollary 5.1 in Ch. 5, Sec. 1.2); hence they are neither in time nor out of it. Nor would a definition of time in terms of the bare individuals of Ch. 1 do, because no state space and a fortiori no event space can be defined for them. We need the concept of a changing thing, hence that of state, if we are to define duration as a sort of distance between states.

Our starting point is then the concept of a lawful state space for a concrete object, or thing, of any kind. (Recall Ch. 3, Sec. 2.5.) In Ch. 5, Sec. 2.5, we saw that the states of a thing can be ordered by reference to those of a standard thing or reference frame, such as a clock mounted on a ruler. That was an *extrinsic* order, i.e. one induced on the thing states *ab extrinseco* by the natural or intrinsic ordering of the reference states. But actually any thing can serve as a clock even though only a few things are good clocks. In other words, the states of every thing are ordered. Because this order is intrinsic and independent of the kind of thing, we are justified in identifying it with temporal order. More explicitly, we assume the *intrinsic* time order of thing states:

POSTULATE 6.7 For every basic thing $x \in B$ and every lawful state space $S_{\mathbb{L}}(x)$ for x, there is exactly one ordering relation $\leqslant$, connected in $S_{\mathbb{L}}(x)$, such that, for any $s, s' \in S_{\mathbb{L}}(x)$,

$$s \leqslant s' \quad \text{iff} \quad s' = g(s), \quad \text{where} \quad g\colon S_{\mathbb{L}}(x) \to S_{\mathbb{L}}(x)$$

is a really possible (lawful) transformation of $S_{\mathbb{L}}(x)$, i.e. one representing a lawful change of x.

This assumption acquires a precise ontological sense by virtue of its semantic partner:

POSTULATE 6.8 The set $S_{\mathbb{L}}(x)$ of states of any basic $x \in B$ is *ordered temporally* by $\leqslant$. Furthermore for any $s, s' \in S_{\mathbb{L}}(x)$

(i) s *precedes temporally* s' iff $s \leqslant s'$;

(ii) s is *simultaneous with* s' iff $s \leqslant s'$ and $s' \leqslant s$ (i.e. $s \sim s'$).

The basic temporal relations have been defined only for the states of a single thing. So far we do not have any concept of temporal relations between states of different (in particular distant) things. It will also be noted that the preceding elucidations depend upon the notion of lawful change but not upon that of causation. Our postulates allow for temporal relations in a stochastic universe, i.e. one in which every single law is stochastic. Causality is sufficient for temporality and it may be necessary to *ascertain* which events precede which, but it is not necessary for the relations of before and after to *obtain*. The relation between precedence and causation is this:

For all events e and e', if e causes e' then e precedes e'.

But the converse is false. Hence precedence in time is not definable in terms of causation, nor is the converse move possible. (See Bunge, 1959.) Therefore our theory of time, though relational, is not causal – unlike the theories of Robb (1914, 1921, 1936) and Reichenbach (1928, 1956). Ours is a nomological rather than causal theory of time.

The equivalence relation $\sim$ introduced by Postulate 6.8(ii) effects a partition of $S_{\mathbb{L}}(x)$ into the states which are simultaneous and those which are not. The latter are related either by the *before* relation $<$ or by the *after* relation $>$. However, if we were to consider a rather bulky thing rather than a basic component (or member of B), there might appear pairs of states which are neither simultaneous nor related by $<$ or by $>$. Such would be the states of components of a thing not connectible by light waves. This is one reason for defining temporal relations on B rather than Θ.

We can now clarify the notions of present, past, and future of a basic thing:

DEFINITION 6.8 Let $S_{\mathbb{L}}(x)$ be a state space for a basic thing $x \in B$, and call $s_0 \in S_{\mathbb{L}}(x)$, a distinguished state of x, the *origin* or *zero* state. Then

(i) the *present* of $x =_{df} \{s \in S_{\mathbb{L}}(x) | s \sim s_0\}$;
(ii) the *past* of $x =_{df} \{s \in S_{\mathbb{L}}(x) | s < s_0\}$;
(iii) the *future* of $x =_{df} \{s \in S_{\mathbb{L}}(x) | s > s_0\}$.

Obviously, we are free to choose the origin s_0 and may call it *now* if we want to – provided we do not insist that this word designates the egocentric "now". Indeed, s_0 may be chosen to be the beginner of a process in the thing concerned, and all interested observers may agree on this choice for the sake of convenience. This, *pace* Hugo Bergmann (1929) and Grünbaum (1967), does not entail that becoming is unreal, hence temporal order conventional. Likewise the relativity of the splitting of states into present, past and future, to a given thing (in the above the very thing whose states are being partitioned) implies no subjectivity. Relativity and parochialism – as shown by the special choices of thing and of initial state – are one thing, subjectivity another.

The relativity of temporal orderings to things cannot be remedied: the two theories of relativity – the special and the general – have taught us that every thing has its proper time, so that there is no universal time. (There would be one if the universe were to pulsate as a whole – but this it does not.) In other words *there is no absolute or thing-free time.* This is a generalization of the statement in relativistic physics, that time is relative to some reference frame or other. We feel justified in making that generalization because, as noted before, almost anything can be regarded as a clock since all things are assumed to tick in some way or other. Of course some things tick more regularly than others and therefore make better clocks. But this is a matter for metrologists, astronomers, and atomic physicists, not for ontologists. From the point of view of ontological principles any thing the states of which are ordered in a certain way qualifies as a clock, whether or not it is actually used as such. The order in question is the strict partial order, i.e. asymmetric and transitive, so that no two states are simultaneous. Therefore we adopt

DEFINITION 6.9 A thing $f \in \Theta$ is a *potential clock* iff every one of its state spaces $S_{\mathbb{L}}(f)$ is strictly partially ordered.

DEFINITION 6.10 Let f be a potential clock with state space $S_{\mathbb{L}}(f) = T_f$. Then

(i) T_f is called the *time span* of f;
(ii) any interval of T_f is called an *f-moment*; and
(iii) any member of T_f is called an *f-instant.*

Each thing, and in particular each clock, ticks its own way. However, things are not monads: recall Postulate 5.10 in Ch. 5, Sec. 4.1. That is, every thing is connected with other (though by no means with all) things. And some such connections or couplings are of the physical information kind, so they can establish relations among the local temporal orderings. (For example, if two bodies move with constant velocity relative to one another then their time coordinates are related by a Lorentz transformation.) This gives us a revised notion of temporal universality: time orders are universal not because they are the same for all things – which they are not – but because they are mutually translatable within bounds. The kind of translation is specified by the following convention and the subsequent axiom.

DEFINITION 6.11 Let $f \in \Theta$ be a potential clock [Definition 6.9] with state space $S_{\mathbb{L}}(f)$ and $x \in \Theta$ any thing connectible with f, with state space $S_{\mathbb{L}}(x)$. Then

(i) an *f-time assignment* to the states of x is an injection

$$\varphi_f: S_{\mathbb{L}}(f) \rightarrow S_{\mathbb{L}}(x);$$

(ii) an f-time assignment φ_f is *faithful* iff, for any $t, t' \in S_{\mathbb{L}}(f)$,

$$t \leqslant t' \Rightarrow \varphi_f(t) \leqslant \varphi_f(t').$$

The next axiom ensures that time assignments are always possible, so that every potential clock can be "utilized" by any thing connectible with it:

POSTULATE 6.9 Let $x \in \Theta$ be any thing. Then there exists a potential clock $f \in \Theta$ and a faithful time assignment φ_f: $S_{\mathbb{L}}(f) \to S_{\mathbb{L}}(x)$.

In other words, the states of any thing can be ordered by the (reference) states of a potential clock connectible with the given thing.

So much for time order and its universal translatability (universal within bounds). We proceed now to build the concept of duration.

3.3. *Duration*

Relativistic physics teaches that durations are frame-dependent. Any given process in a thing has as many durations as there are (inequivalent) potential clocks connectible with the given thing. (And the shortest of all such durations is the one relative to the thing itself or to a frame attached to it. That is, the shortest duration is the so-called *proper* duration, i.e. the one relative to the thing undergoing the process of interest.) Therefore our ontological theory of time must take into account the relativity of duration.

We assign each event or process in a thing a collection of contiguous reference states, i.e. an interval of the time span of a clock. (See Figure 6.5.) The assignment is made by

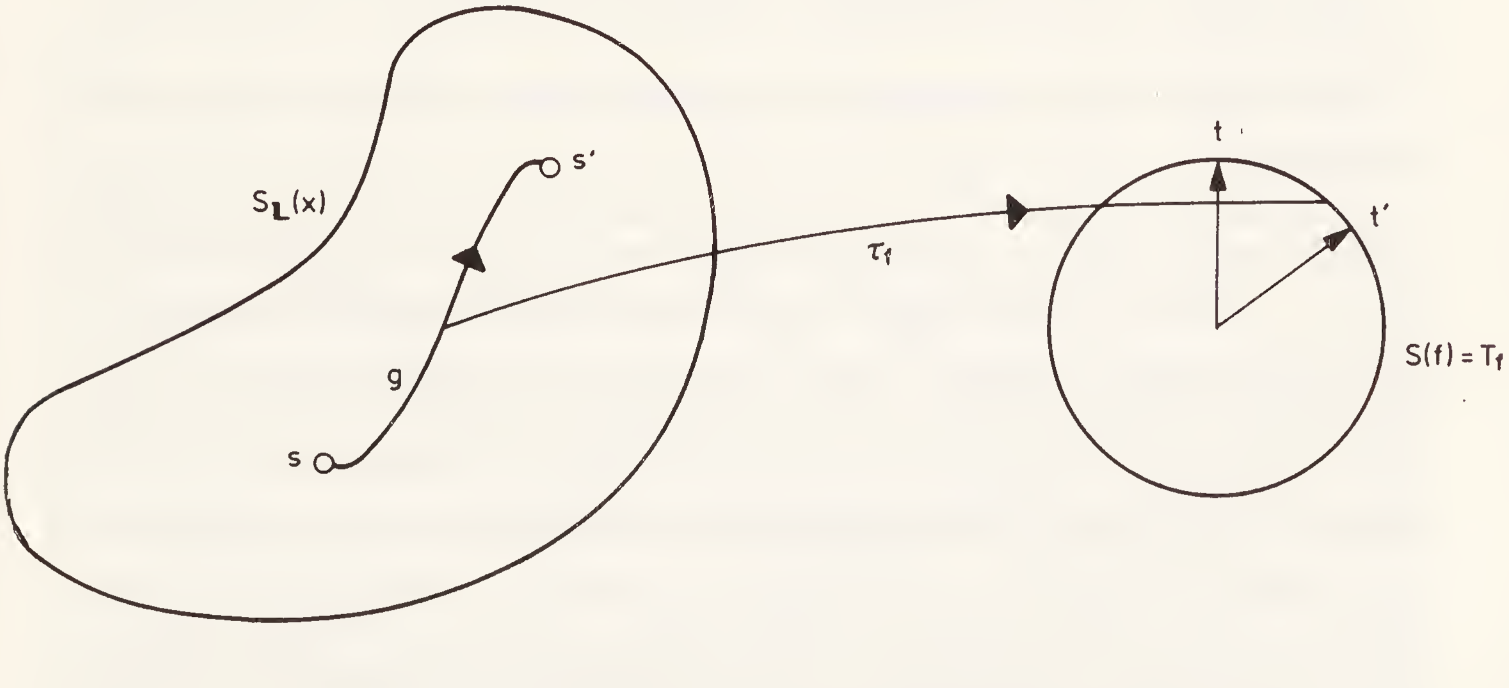


Fig. 6.5. The process $\langle s, s', g\rangle \in E_{\mathbb{L}}(x) \subseteq S_{\mathbb{L}}(x) \times S_{\mathbb{L}}(x)$ in thing x is assigned the interval $[t, t'] \subset T_f$ of reference states.

DEFINITION 6.12 Let T_f be the time span of a clock f and let $E_{\mathbb{L}}(x) \subseteq S_{\mathbb{L}}(x) \times S_{\mathbb{L}}(x)$ be an event space for a thing x connectible with f. The *duration relative to f*, or *f-duration*, is the function

$$\tau_f \colon E_{\mathbb{L}}(x) \to 2^{T_f}$$

from events to intervals of reference states such that, for any events $e = \langle s, s', g \rangle$, $e' = \langle s'', s''', h \rangle \in E_{\mathbb{L}}(x)$, where g and h are lawful transformations of $S_{\mathbb{L}}(x)$,

(i) if $s \sim s'$ then $\tau_f(e) = \varnothing$;
(ii) if $s < s'$ or $s' > s$ then $\tau_f(e) \supsetneq \varnothing$;
(iii) if e and e' compose to form $e'' = \langle s, s''', g \circ h \rangle \in E_{\mathbb{L}}(x)$, then

$$\tau_f(e) \cup \tau_f(e') = \tau_f(e'')$$

$\tau_f(e)$ is called the *duration* of e relative to f.

The duration function enables us to define in a precise (yet still nonquantitative) manner a number of notions in wide usage:

DEFINITION 6.13 Let τ_f be the duration function associated with a clock-thing pair $\langle f, x \rangle$. Then for any two events $e, e' \in E_{\mathbb{L}}(x)$ occurring in thing x,

(i) e is *instantaneous* (or a *point event*) iff $\tau_f(e) = \varnothing$;
(ii) e is *temporally extended* (or *takes time*) iff e is not instantaneous;
(iii) *e lasts just as long as e'* relative to $f =_{df} \tau_f(e) = \tau_f(e')$;
(iv) *e lasts longer than e'* relative to $f =_{df} \tau_f(e) \supset \tau_f(e')$;
(v) e and e' are *temporally contiguous* relative to $f =_{df} \tau_f(e) \cap \tau_f(e') \neq \varnothing$;
(vi) e and e' are *temporally detached* iff they are not temporally contiguous.

An identical transition $s \mapsto s$, or nonevent, is of course trivially instantaneous. But it may also happen that $\tau_f(s, s', g) = \varnothing$ even if $g(s) \neq s$. That is, there might exist instantaneous proper events, i.e. genuine changes taking no time. Whether such changes exist we do not know and we need not know for building our theory of time, because the values of the duration function are intervals between $\varnothing$ and T. Hence we do not face the difficulty met with by alternative theories of time – such as Russell's and Whitehead's – based on the assumption that all events are analyzable into point (instantaneous) events. (This is a difficulty because, according to the quantum theories, there are temporally extended elementary events.)

We have now all we need to state in a precise fashion the gist of the relational theory of time:

DEFINITION 6.14 Let x be a thing and $E_{\mathbb{L}}(x)$ an event space for x. Further, let τ_x be the duration function relative to x. [I.e., x is now its own reference frame or clock.] Then the ordered couple $\langle E_{\mathbb{L}}(x), \tau_x \rangle$ is called the *local time* in (or relative to) x.

If $E_{\mathbb{L}}(x)$ were to shrink to nought, i.e. if x were immutable, the duration function would lose its support and there would be no time left. In other words we infer

COROLLARY 6.3 There is no time where there are no changing things.

Hence we are justified in calling a thing *timeless*, or out of time, just in case it is unchanging. But by Postulate 5.11 in Ch. 5, Sec. 4.1 there are no immutable things. Hence there are no timeless things. Therefore if anything looks timeless then it has been improperly investigated.

We emphasize that, since $E_{\mathbb{L}}(x)$ is the set of lawful events in thing x, our time concept hinges on the notions of thing, change, and law. In a chaotic (i.e. lawless) thing there would be no time. (Remember the distinction between stochastic orderliness or regularity, on the one hand, and chaos on the other: cf. Ch. 4, Sec. 6.3.) If we can discern temporal relations even in an arbitrary set of events, such as those composing a slice of our experiential stream, it is only relative to some set of regular events such as those of our biological clock.

Note also that Definition 6.14 allows for as many times as there are things. The assumption that all these times are the same, i.e. that the pace of events is the same relative to all things, is the hypothesis of *universal time*. We do not adopt this assumption because we wish our metaphysics to be compatible with physics, and the latter admits only local times. However, a relational theory of time including the universal time hypothesis would still be rela*tional* though not rela*tivistic*.

So far our concepts of time are qualitative. We can easily metrize the concept of duration provided we borrow from the foundations of physics the concept of a chronometric scale-*cum*-unit system. Indeed, the precise measure of a duration depends upon the way the successive stages of a process are assigned numbers, i.e. on the chronometric system. Moreover since time units – like all units – are conventional, they can be generated *ad libitum* to form a whole set U_t. Reckoned on a certain time unit $u \in U_t$, the duration of an event $e = \langle s, s', g \rangle$ in a thing x, relative to a clock f, can be taken to be the value, at $\langle e, u\rangle$, of a certain function T_f to be defined presently, i.e. $t = T_f(e, u)$. This function is characterized by

DEFINITION 6.15 Let f be a clock and x a thing, possibly the same as f, with event space $E_{\mathbb{L}}(x)$. Further, let τ_f be the duration function for the given pair $\langle f, x\rangle$. Then the corresponding *metric duration* is the mapping

$$t_f: E_{\mathbb{L}}(x) \times U_t \to \mathbb{R}$$

satisfying the conditions that, for any fixed unit $u \in U_t$ and any two events $e = \langle s, s', g\rangle$, $e' = \langle s'', s''', h\rangle \in E_{\mathbb{L}}(x)$,

(i) $t_f(e, u) \geqslant 0$ iff $\tau_f(e) \supseteq \varnothing$;

(ii) if e and e' compose to form $e'' = \langle s, s''', g \circ h\rangle \in E_{\mathbb{L}}(x)$, then

$$t_f(e, u) + t_f(e', u) = t_f(e'', u).$$

Clearly, metric durations are frame-dependent, which is in keeping with relativistic physics. Note also that clause (i) does not exclude negative durations but states only that, if state s precedes state s', then the duration of this event is non-negative.

Clearly, for every $u \in U_t$, the structure $\langle E_{\mathbb{L}}(x), t_f\rangle$ is a metric space. This justifies our calling $t_f(e, u)$ the *metric duration* of event e, relative to frame f, in units u.

We derive a couple of immediate consequences. First

COROLLARY 6.4 The duration of a point event is nil: for any frame f, If $\tau_f(e) = \varnothing$ then $t_f(e, u) = 0$ for any $u \in U_t$.

Proof By clause (ii) of Definition 6.15 applied to the case of identical change or improper event. In fact in this case $2t_f(e, u) = t_f(e, u)$, whence $t_f(e, u) = 0$.

COROLLARY 6.5 The metric duration is an oriented interval [or directed set]: for any frame f, any unit $u \in U_t$ and all states $s, s' \in S_{\mathbb{L}}(x)$ forming the conceivable events $e = \langle s, s', g\rangle$ and $e' = \langle s', s, g^{-1}\rangle$,

$$t_f(\langle s, s', g\rangle, u) = -t_f(\langle s', s, g^{-1}\rangle, u).$$

Proof By clause (ii) and Corollary 6.4 upon setting $s''' = s$.

The last consequence expresses the so-called *anisotropy* or *asymmetry* of time, about which so much nonsense has been written. There is no mystery in it and it has nothing to do with irreversible processes.

Indeed, Corollary 6.5 is just a consequence of a convention, namely Definition 6.15, which assigns positive metric durations to changes occurring in the natural order and negative durations to their conceptual inversions. (We might as well have chosen the backwards reckoning.) More on this subject in Sec. 5.1.

4. SPACETIME

4.1. *Spacetime, the Basic Network of Events*

Since Einstein's 1905 revolution, space and time have ceased to be regarded as independent: they are now treated as two aspects of a single four-manifold called *spacetime.* The separate relational theories of space and time should accordingly be replaced by a unified theory of spacetime. According to our relational point of view this formal unification is not a mere mathematical trick but formulates the idea that spacetime is generated by changing things. Before proceeding to the mathematical development let us give an intuitive idea.

Imagine a miniature universe composed of just two particles initially superposed at the origin (see Figure 6.6). All of a sudden one of the particles, call it *p*, starts moving while the other remains at 0. (Actually since there is no third thing that can function as a reference frame they just move relative to one another.) As *p* moves the edge of

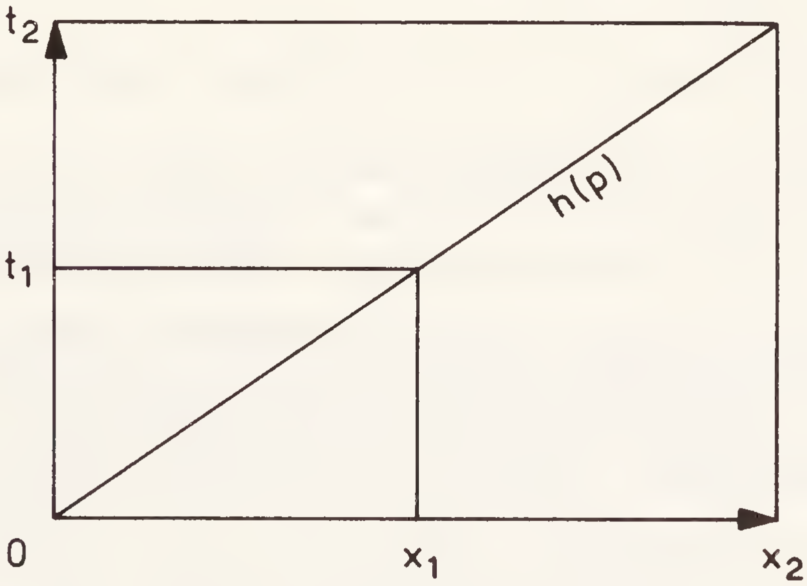


Fig. 6.6. The edge of spacetime advances as particle *p* moves away from the particle located at 0.

spacetime advances. When p reaches position x_1, spacetime covers the interior rectangle. This region expands up to the outer rectangle as p reaches position x_2. If p stops here there is no further displacement of the spacetime edge, and the outer rectangle remains frozen: the system has a past but no future.

We shall construct our view of spacetime on the basis of the concepts of event and of spatiotemporal separation between events occurring in things that may be different. Spacetime will then be defined as a certain set of events together with their spatiotemporal separations. However, this new concept of separation will be based on the concepts of thing separation (Definition 6.1) and of event duration (Definition 6.12). This strategy will allow us to advance rather quickly.

We start by defining the spatial distance between events as the separation between the things where they occur:

DEFINITION 6.16 Let $x, y \in B$ be two basic things, possibly identical, with event spaces $E_{\mathbb{L}}(x)$ and $E_{\mathbb{L}}(y)$ respectively relative to a common frame f. Moreover assume that all three things are connectible. Then the *spatial separation* between the events in x and those in y equals the separation between the corresponding things. In symbols,

$$\delta_{sf}: E_{\mathbb{L}}(x) \times E_{\mathbb{L}}(y) \to 2^B$$

such that

$$\delta_{sf}(e, e') = \sigma(x, y) \quad \text{for} \quad e \in E_{\mathbb{L}}(x) \quad \text{and} \quad e' \in E_{\mathbb{L}}(y).$$

If the things are not connectible – e.g. by signals of some kind – then the inter-event separation is not defined. And if the things coincide then the events are not spatially separated. Indeed, by Theorem 6.1(i), if $x = y$ then $\delta_{sf}(e, e') = \sigma(x, x) = \{x\}$. Note finally that the index f occurring in the event separation function is missing in the thing separation function. This is because, whereas events are relative to frames, their carriers (things) are not. Think of an emitter moving towards a stationary receiver – moving, that is, relative to a certain reference frame. The wavelength of the emitted signal differs from that of the received signal (Doppler effect). On the other hand the set of things that can interpose between the two bodies is frame-invariant.

Next we define the temporal distance between two events as the lapse between the end state of the first and the beginning state of the second. (We are presupposing here Postulate 5.8 in Ch. 5, Sec. 3.2, i.e. that every change has a beginning and an end.) More exactly, we make

DEFINITION 6.17 Let $x, y \in B$ be two basic things, possibly identical, with event spaces $E_{\mathbb{L}}(x)$ and $E_{\mathbb{L}}(y)$ respectively relative to a common frame f with states in $S(f)$. Moreover suppose that all three things are mutually connectible. Then the *temporal separation* between the events in x and those in y is the function

$$\delta_{tf}: E_{\mathbb{L}}(x) \times E_{\mathbb{L}}(y) \to 2^{S(f)}$$

such that, if $e \in E_{\mathbb{L}}(x)$ and $e' \in E_{\mathbb{L}}(y)$,

$$\delta_{tf}(e, e') = \tau_f(\text{end state of } e, \text{ beginning state of } e').$$

Now, the two separation functions δ_{sf} and δ_{tf} can be regarded as so many projections of a third function $\Delta_f = \delta_{sf} \times \delta_{tf}$ determined by the former in accordance with the following diagram:

$$\begin{tikzcd}
& E_{\mathbb{L}}(x) \times E_{\mathbb{L}}(y) \arrow[dl, "\delta_{sf}"'] \arrow[d, "\Delta_f"] \arrow[dr, "\delta_{tf}"] & \\
2^B & 2^B \times 2^{S(f)} \arrow[l] \arrow[r] & 2^{S(f)}
\end{tikzcd}$$

More explicitly, we propose

DEFINITION 6.18 Let $x, y \in B$ be two basic things with event spaces $E_{\mathbb{L}}(x)$ and $E_{\mathbb{L}}(y)$ respectively relative to a common frame f with states in $S(f)$. Moreover assume that all three things are mutually connectible. Then the function

$$\Delta_f = \delta_{sf} \times \delta_{tf}: E_{\mathbb{L}}(x) \times E_{\mathbb{L}}(y) \to 2^B \times 2^{S(f)}$$

such that

$$\Delta_f(e, e') = \langle \delta_{sf}(e, e'), \delta_{tf}(e, e') \rangle \quad \text{for} \quad e \in E_{\mathbb{L}}(x), e' \in E_{\mathbb{L}}(y),$$

is called the *spatiotemporal separation* (or *interval*).

We have now all we need to define a philosopher's relational concept of (local) spacetime:

DEFINITION 6.19 Let $B_0 \subset B$ be a set of basic things and call f a common reference frame. Furthermore suppose that all these things, i.e. $B_0 \cup \{f\}$, are mutually connectible. Finally, call E_f the event space, relative to f of the aggregation $[B_0]$ of basic things, and Δ_f the corresponding spatiotemporal separation. Then the ordered pair $\langle E_f, \Delta_f \rangle$ is called the *f-spacetime* (or *local spacetime of frame f*).

In physics a far more sophisticated structure is defined, namely a connected four-manifold M_f^4. We shall not construct M_f^4 out of the event f-spacetime. But this can be done along lines similar to those followed in Sec. 2 for building the physicist's space. In other words it is possible to map the events connectible with a frame f, and their mutual separations, on to a local spacetime chart M_f^4. And once such a chart for an arbitrary frame f has been defined, one can compose the atlas formed by all such local charts – i.e. the global or cosmic spacetime M^4.

So much for our outline of a relational theory of spacetime. Such a theory is not only relational but also compatible with relativistic physics, in that (*a*) it assumes the structure of spacetime to depend upon its furniture, and (*b*) it does not postulate a global structure. However, the theory is not *relativistic*: it does not include any of the special laws characterizing the various relativistic theories, such as for example the frame independence of the velocity of light, or the equations of the gravitational field. The relational theory of spacetime sketched above is just a component of the background of any general-relativistic theory – if one cares to add such an ontological background. Physicists usually don't: they are in the habit of postulating the four-manifold without inquiring into its roots in events. (Surely they call *events* the points of the manifold, but this is not done consistently, for such a definition presupposes a bijection between events and points in spacetime, and actually there is no such bijection, i.e. they do not mean that there is something going on at every point in spacetime.) Conversely, a relational theory of space and time may not be consistent with relativistic physics. Thus Robb's original theory (Robb, 1914), based on a set of point-like events and an absolute (frame-independent) relation of precedence, was relational but at variance with relativistic physics. In sum the theory of spacetime outlined in this subsection is relational and compatible with relativity but not relativistic. Only relativity is relativistic; and it is also relational provided it is explicitly based on a relational theory of spacetime.

4.2. *Position in Spacetime*

It is likely that events, however elementary, are not pointlike but take up nonvanishing regions of spacetime. That is, the event space E_f associated with a frame f ought to be mapped on the power set of the four-manifold M_f^4 rather than on M_f^4 itself. However, we may pretend, for the sake of simplicity, that events are pointlike, so that each event $e \in E_f$ can be assigned a point p in the corresponding local chart M_f^4. We shall make this simplifying assumption here because we are concerned only with clarifying certain ideas. The refinement is left to the reader.

We assume then that the event spacetime E_f can be mapped on a chart M_f^4. Now, by definition of a manifold, the latter can in turn be mapped locally on a set of ordered quadruples of real numbers. Consequently by composing the two maps we can assign numbers to events. This can be called a spatiotemporal coordinatization of events. The gist of it is shown in Figure 6.7, where 'N_f' designates an open subset of M_f^4.

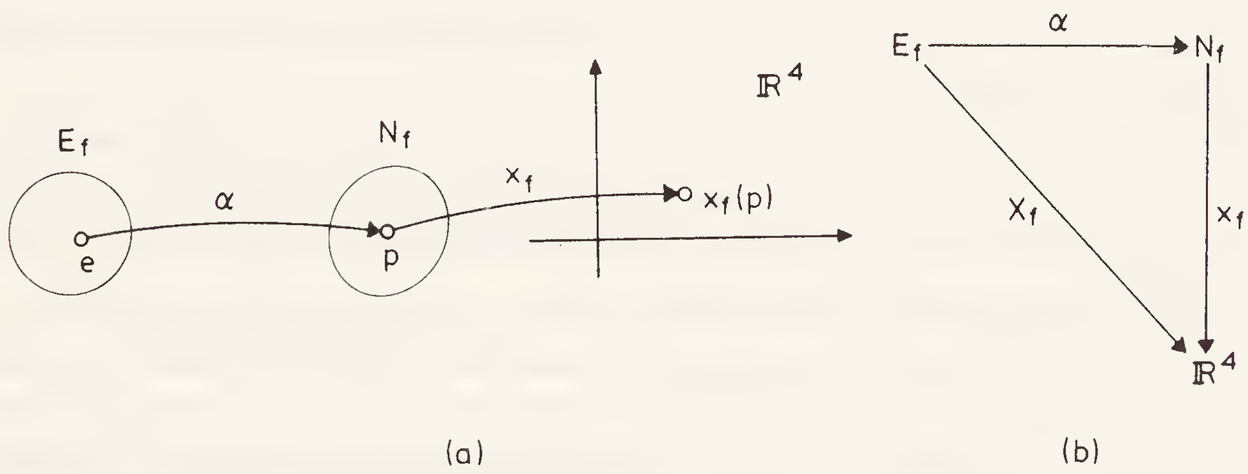


Fig. 6.7. (*a*) The mapping of the image N_f of E_f onto numbers and (*b*) the composition of the maps α and x_f to form the map X_f. The latter is a physical coordinate while x_f is a geometric coordinate. Whereas α is injective, x_f is bijective. Hence although every event is paired off to a quadruple of reals, the converse is false: there might be points in spacetime where nothing happens. (This is impossible according to our ontology, but scientific theories are partial and most of them do not assume that something is going on at every point in spacetime.)

In other words, for every reference frame f there is a coordinate system $\langle M_f^4, x_f \rangle$ representing f, where x_f, the local coordinate function, is a homeomorphism from an open subset N_f of M_f^4 into the cartesian space $\mathbb{R}^4$. This homeomorphism assigns each *point* $p \in N_f$ a quadruple of real numbers called the *coordinate values* of p relative to f:

$$x_f(p) = \langle x_1(p), x_2(p), x_3(p), x_4(p)\rangle, \quad p \in N_f \subset M_f^4.$$

(The choice of either of the four coordinate functions as the time coordinate is conventional, but one of them has got to be interpreted as the time coordinate.)

On the other hand the value of the composition of x_f and α, i.e.

$$X_f(e) = (x_f \circ \alpha)(e), \quad \text{where} \quad e \in E_f,$$

equals the spatiotemporal location of the *event e* relative to f. Note the difference – seldom recognized – between the geometrical coordinate functions x_f and the physical coordinate functions X_f. (The plural is apposite because for each f there are four coordinate functions of each kind.) The former locate *constructs*, the latter locate *events* represented by those constructs. Because α is not onto, X_f is not onto either, i.e. there are points and quadruples of reals that represent no events at all. But because α and x_f (hence also X_f) are functions, *every event is somewhere in spacetime* (*relative to a frame*). And because every event is a change in some thing, it follows that *every thing is at some place or other* (*relative to a frame*).

Nowadays we take it for granted that, even if space (and time) were absolute – i.e. self-existent –, position (in space and time) is not. But the idea that places are relative to frames has not been easy to come by. In all archaic and ancient cosmologies place was regarded as absolute and unique, every thing was supposed to occupy a definite place in space, and every event to occur at a moment of time likewise independent of any frame. These ideas are found not only in Aristotle but as recently as in Strawson (1959, pp. 22, 25–26). The credit for exactifying the notion of place relative to a coordinate system usually goes to Descartes. Galilei and his followers introduced the notion of a physical reference frame, which they dimly saw as representable by, though not identical with, a coordinate system. By the same token they showed that things and events have no absolute places – or, equivalently, that every single thing or event has as many places as there are reference frames, i.e. in principle infinitely many. These developments, stemming from Descartes and Galilei, disposed then of the principle that every thing and every event have or occupy a unique hence absolute place. In our own century Einstein crowned that development by demolishing the axiom that every fact happens relative to any reference frame (not just relative to the connectible frames) and consequently that it happens at a single absolute time irrespective of any frames.

Other developments that completed the demise of the *One fact-one place in space and time* principle were the following. First, Newton hypothesized that gravity can act everywhere, whence it has no unique or natural place. Second, Faraday showed that electricity and magnetism can likewise extend almost anywhere. Third, different bosons (e.g. photons) have been shown to be able to occupy the same place. Fourth, quantum mechanical things – whether particles or fields – seem to occupy no clear cut spatial regions, i.e. they have no sharp boundaries. These discoveries have refuted the old principles concerning place and have even diminished the importance of the category of place. Parallel developments in epistemology have devalued the notions of here and there, now and then. These have proved to be egocentric particulars (Russell, 1942) rather than objective items. That is, they are meaningful only in relation to a very particular reference frame, namely the self. Remove all subjects and no trace is left of here and there, now and then. Spacetime, on the other hand, is supposed to outlive the hecatomb for, *pace* Kant, it is not phenomenal.

Place in spacetime is then viewed as being relative to some reference frame or other. Being relative, location in spacetime is not an intrinsic property of things and events. And, being neither intrinsic nor causally effective, place in spacetime is not essential *per se*. In other words, coordinate values are rather arbitrary labels. What may be of importance is not the place that the mapping α assigns to an event nor consequently the relative positions of points and regions in the spacetime manifold, but rather the connections or bonds among things. In particular, it is irrelevant whether thing a is near thing b unless such nearness makes it possible for a to act upon b or conversely.

To sum up. Location in spacetime is both relative to a frame and objective. It is epistemologically important because it enables us to individuate or identify things and events that, aside from their distinct spatiotemporal location, might be equal (i.e. "only numerically distinct", as it used to be said). But spatiotemporal location is arbitrary to the extent that the choice of reference frame is arbitrary within broad bounds. Hence it is ontologically of secondary importance. So much so that the basic laws are invariant under changes in spatiotemporal position – which is a way of saying that coordinates are artifacts. But this is a matter for the next section.

4.3. *Change in Spacetime*

Change in general can be described without the assistance of any concepts of space and time: see Ch. 5. But specific kinds of change, such as change of place (relative to some frame) require of course some concepts of space and time. The very first thing to do in order to describe the motion (Aristotle's local change) of a thing is to model the latter either as a piece of space or as a chunk of spacetime (see Figure 6.8).

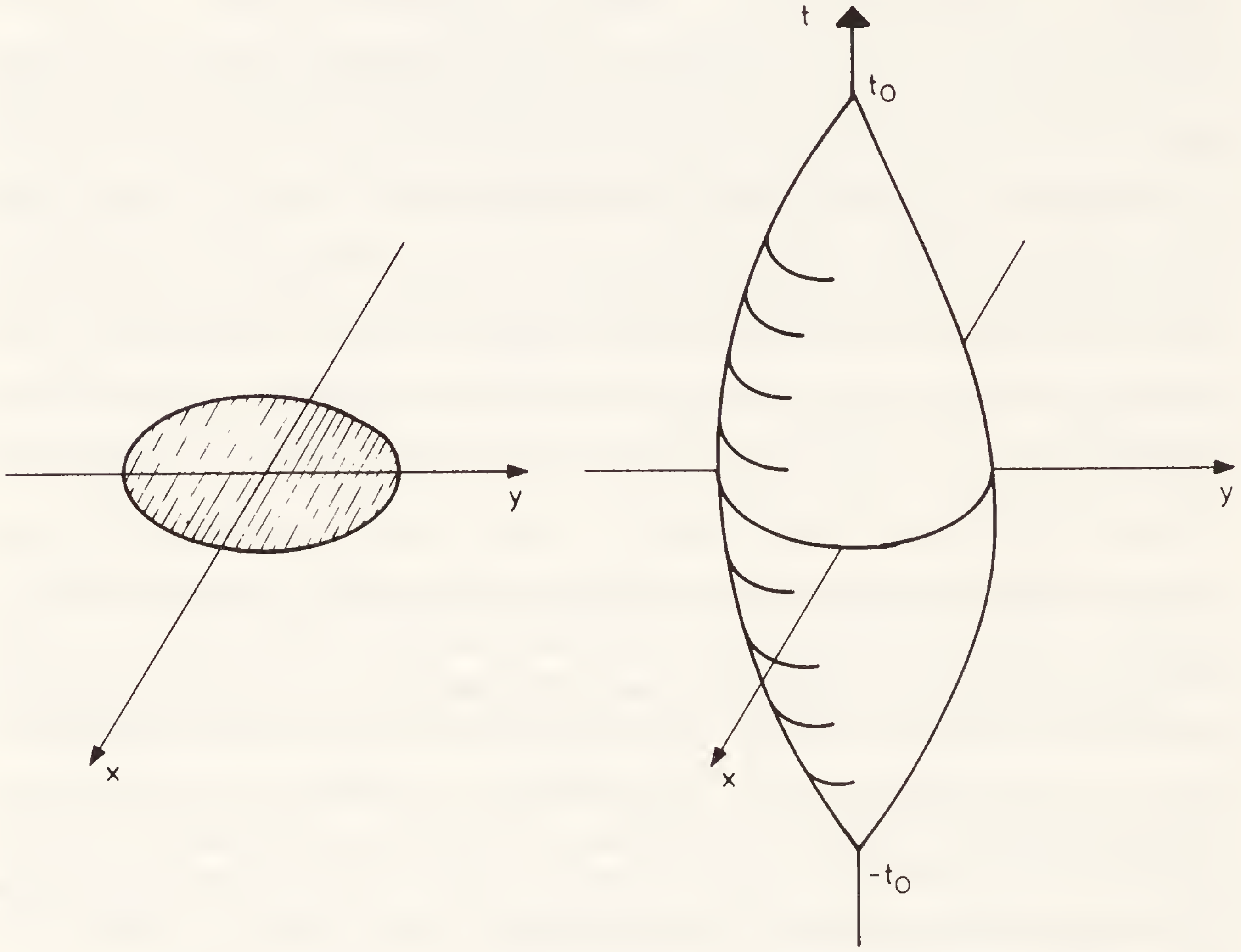


Fig. 6.8. Growth and decline of a disk-shaped thing. (*a*) Spatial representation of thing at state of greatest bulk. (*b*) Spatiotemporal representation of history of thing from birth at $-t_0$ to extinction at t_0. The thing as a whole does not move relative to the chosen reference frame.

Motion is then described as the change of place, in the course of time, of some or all of the points making up the thing – points which, unlike those of empty space, are located with the help of physical coordinate functions X_f (Sec. 4.2).

In building a model of a moving thing we start from Definition 3.6, in Ch. 3, Sec. 1.4, of a functional schema. A functional schema of a thing X

is a model thing X_m consisting in an ordered couple $X_m = \langle M, \mathbb{F}\rangle$, where M is the domain on which the state function $\mathbb{F}$ is defined. The latter is usually frame-dependent and it consists in an n-tuple of functions, usually real valued ones. We obtain a mechanical model if we equate M with space, or with spacetime, and some of the components of $\mathbb{F}$ with the physical coordinates of the thing (relative to the given frame), in the sense of Sec. 4.2. The values of these coordinate functions differ from zero only at the places (in space or in spacetime) at which the thing is. More precisely, we can make

DEFINITION 6.20 Let $X_m = \langle M, \mathbb{F}\rangle$ be a functional model of a thing X relative to a reference frame f [that occurs in the very construction of $\mathbb{F}$]. This model is a *mechanical model* just in case

(i) $M = N_f \subset M_f^4 =$ the spacetime chart attached to f;

(ii) some components of the state function $\mathbb{F}$ are physical coordinates serving to locate the parts of the thing:

$$X_f\colon N_f \to \mathbb{R}^4 \quad \text{or} \quad X_f\colon N_f^3 \times T_f \to \mathbb{R}^4$$

where N_f^3 is a three-dimensional projection of N_f.

In most theories the second choice of position coordinates is available, so that a general value of the position vector would be

$$X_f(x_f, t) = \langle x, y, z\rangle \in \mathbb{R}^3.$$

Note that the physical coordinate depends upon the geometric coordinate. The values of the two functions coincide only where the thing is. It is only in very special theories, such as classical particle mechanics, that X_f depends not on x_f but only on t.

A special kind of mechanical model is that of kinematical model: here all of the state functions are physical coordinates. That is, we have

DEFINITION 6.21 Let $X_m = \langle N_f, \mathbb{F}\rangle$, with $N_f \subset M_f^4$ be a mechanical model of a thing X. This model is a *kinematical model* iff $\mathbb{F} = X_f$.

The concept of motion, or mechanical change, may be regarded as a specification of the general concept of history introduced by Definition 5.27 in Ch. 5, Sec. 3.2, as shown by

DEFINITION 6.22 Let $X_m = \langle N_f, \mathbb{F}\rangle$, with $N_f \subset M_f^4$, be a mechanical model for a thing X relative to a frame f, and call X_f the position coordinate functions in $\mathbb{F}$. Further, call T_f the time projection of the state space $S(f)$ of frame f. Then a *motion* (or *mechanical change*) *of X relative to f* is any nonempty subset of its kinematical history

$$\mu_f(X) = \{\langle t, X_f(t, x_f)\rangle \in \mathbb{R}^5 | t \in T_f\}.$$

Spacetime is so important because it is an inescapable part of the state space of any moving thing. (The two seldom coincide, for change of place often brings about changes in other properties; or conversely. It is only in extremely idealized models, such as that of the point particle, that the history coincides with the kinematical history.) Moreover whenever there is change of any kind there is also motion, if not of the whole thing at least of some of its parts. (Think of a community established in an old town: although the town does not move relative to the earth, every member of the community moves about a great deal.) This is so important and sweeping a generalization that it deserves being conferred the title of

POSTULATE 6.10 The history of any basic thing, relative to an arbitrary frame, has a nonempty projection onto the spacetime attached to the frame.

This postulate must not be mistaken for the hypothesis that all change is reducible to motion. This other thesis, which characterizes all mechanistic ontologies, is plainly false: just think of the propagation of an electromagnetic wave or of an industrialization process. Surely the things concerned, or at least some parts of them, move in the course of such processes; but they also change in other respects. This remark too deserves being generalized and the generalization exalted to the axiom rank:

POSTULATE 6.11 Whatever changes does so in more than one respect.

The implication for epistemology is clear: no theory concerning a single type of change suffices to account for the multifarious changes of real things.

5. SPATIOTEMPORAL PROPERTIES

5.1. *Does Spacetime have any Properties*?

Spacetime is often said to have mathematical properties. According to our view only mathematical objects can have mathematical properties and, since spacetime is not such an object, it can have no mathematical properties. Not spacetime itself but rather every one of its conceptualizations, in particular the four-manifold M^4, has mathematical properties, such as that of being the carrier of continuous functions. That the distinction between spacetime and its mathematical representations is seldom made in the scientific literature should be no excuse for the philosopher's failure to draw it.

How about physical properties: does spacetime have any? On our view spacetime has no physical properties either. How could spacetime have substantial properties not being a thing but rather a (relational) property of things? Still, it is often stated dogmatically that spacetime does have certain physical properties, in particular the following three. One is that spacetime is the "carrier" of physical fields such as the gravitational field – a function formerly assigned to the ether. This is suggested by the mathematical representation of a physical field as a tensor (or spinor or vector) field over a manifold. But a mode of representation should not be identified with the object represented. Physical fields need no carrier or substrate: they are things on their own – as first recognized by Einstein. Surely we shall continue to speak of, say, the gravitational field at a given place, without however attributing places an autonomous existence nor, a fortiori, physical properties.

Another property commonly attributed to spacetime is that any given event is connectible only with events lying inside its forward light cone (see Figure 6.9). This restricted connectibility is sometimes said to be a

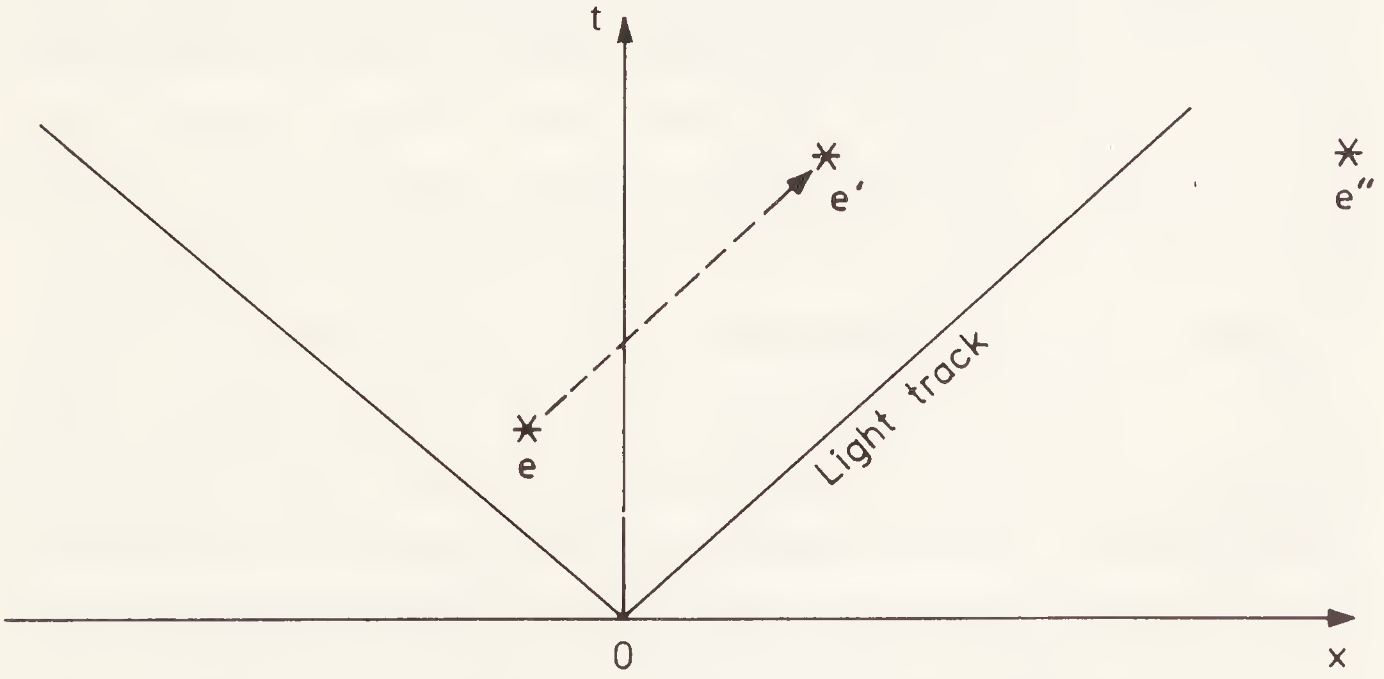


Fig. 6.9. Events e and e' are connectible, but e and e'' are not because there are no (known) superluminal signals.

fundamental property of spacetime. Yet note that (*a*) connectibility is definable as a ternary relation among physical entities, and (*b*) connectibility depends on there being things capable of moving. If superluminal entities (tachyons) were discovered we would cease to regard connectibility within just the light cone as a universal property of things and their changes

A third important instance of the attribution of physical properties to spacetime concerns the vacuum polarization postulated in quantum field theory. Actually this is not a property of spacetime but of the background field, e.g. of the electromagnetic field in its ground state, where it exists and fluctuates but has no quanta (photons).

We generalize the preceding considerations, in line with our relational view of spacetime, and state that the latter *has no properties*, and this because it is itself a property, namely the basic mesh of the sum total of changing things. It is things and their changes, as well as the patterns of either – i.e. the laws – that have spatiotemporal properties. For example a hydrogen atom in its ground state is spherically symmetric, and the development of an embryo is not time symmetric, or rather reversible – but these are properties of the things concerned not of spacetime.

An important spatiotemporal characteristic of a factual item is its absoluteness or the lack of it. Some facts, such as light emission, childbirth and industrial production, are absolute, i.e. independent of any reference frame. Others are not: for example a light detector, such as a photocell or a retina, tuned to respond to a certain wavelength band, will detect certain signals coming from a stationary source but will fail to detect the same radiation if the source moves either towards or away from the sensor. Light absorption depends then upon the reference frame. Needless to say a relative fact is as objective as an absolute one. The circumstance that the detector might belong to a living being, in particular a subjectivist philosopher, does not render the fact subjective.

Likewise while some laws are absolute others are relative to some frame. In particular all the basic laws are invariant under spatial (actually spatiotemporal) translations and rotations. That is, place and spatial orientation make no difference to the basic patterns. (Most laws are also reflection invariant, i.e. they do not distinguish between left and right, or East and West.) In other words, if a basic law obtains in a given region $m \subset M^4$ of spacetime, then it also holds in any other region $m' \subset M^4$. This principle of spatiotemporal invariance (or rather covariance) applies only to the basic laws: the derived laws need not have this property, since they may embody particular circumstances of a spatiotemporal character, such as initial conditions and boundary conditions. Another caution: the preceding principle does not entail that, if there are biological (or sociological) laws in a given region, then the same laws are also in force in any other region of spacetime, even if devoid of organisms and a fortiori of societies. The principle is satisfied vacuously by the regional laws that happen to obtain in a region but are not "exemplified" in another.

5.2. *Time Reversal and Process Reversibility*

The matter of invariance under time reversal deserves a special section because it is often misunderstood. Call π a net representing a linear chain of states of some thing, i.e. a nonbranching process (relative to some reference frame of course). Call $\tilde{\pi}$ the transpose of π, i.e. the same set of states in reverse order. For example, if $\pi = \langle s_1, s_2, s_3 \rangle$ then $\tilde{\pi} = \langle s_3, s_2, s_1 \rangle$. The given net and its transpose are related by the order reversing operator T such that $\tilde{\pi} = T\pi$. (Caution: given π and T, the transpose is uniquely determined. But in general T is not uniquely determined by π and π', i.e. the preceding equation does not define T uniquely. As it so often happens the inverse problem does not have a single solution.) The operator T is usually called the *time reversal* operator. This is a misnomer because T represents the inversion of the order of states not the reversal of time itself (see Figure 6.10). A better name would be that of *process reversal* operator.

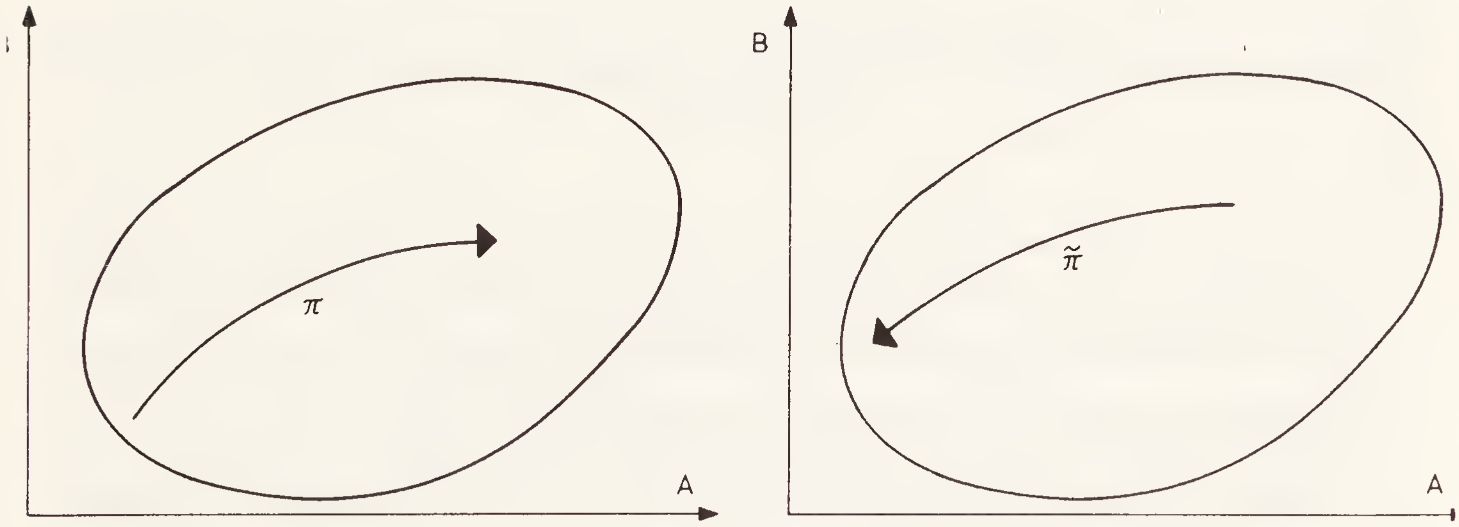


Fig. 6.10. A process π in the state space $S \subset A \times B$ of a thing and the transpose $\tilde{\pi} = T\pi$ of π. T maps the first onto the second.

The relation $\tilde{\pi} = T\pi$ between a process and its transpose can be interpreted in two different ways. A first interpretation is this: if we decide to count time intervals backwards, e.g. by using a clock whose pointers turn counter-clockwise, then we must invert the order of the states: the last state will come (temporally) before all the previous states, as shown in Figure 6.11. An alternative interpretation of the same mathematical operation of time (or rather state order) reversal is this. The transpose $\tilde{\pi}$ of π is not just a nonstandard representation of π but stands for a possible process going forward in time and traversing in reverse order the same stages as π – as when a film is projected backwards. However, this second interpretation of time reversal is legitimate only provided $\tilde{\pi}$ is lawful, which is not the case with every process. When a process happens to have a transpose going forward in time it is called a *reversible* process – otherwise it is said to be *irreversible*. (Recall Definition 5.21 in Ch. 5, Sec. 3.1.)

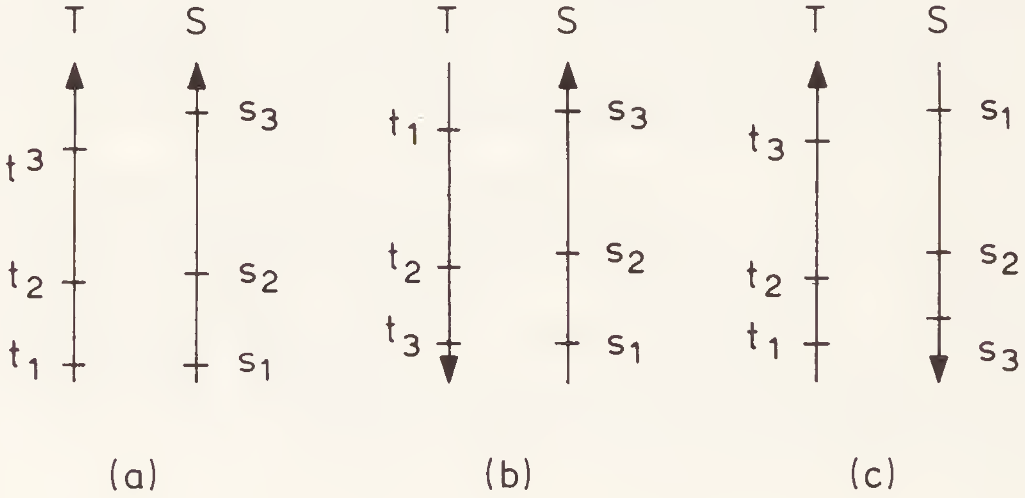


Fig. 6.11. (*a*) The standard instant-state correspondence. (*b*) The queer but possible representation of a process by its transpose. (*c*) The actual inversion of a process in case it is reversible.

If a process is reversible, the time coordinate values of its successive states afford a convenient if misleading representation. By Corollary 6.5 in Sec. 3.3, the metric duration function t_f is antisymmetric (odd) in the sense that, if s and s' belong to the same process, then $t_f(\langle s, s', g\rangle, u) = -t_f(\langle s', s, g^{-1}\rangle, u)$ for every frame f and every scale-cum-unit system u. Calling t the value $t_f(\langle s, s', g\rangle, u)$, as is usual in science, we discover the factual meaning of the inversion of the sign of t: it corresponds to an inversion of the order of appearance of the underlying states. That is, *time reversal represents process reversal*:

$$\textit{forward process } \langle s, s', g \rangle, \quad t_f(\langle s, s', g \rangle, u) = t$$

$$\textit{reverse process } \langle s', s, g^{-1} \rangle, \quad t_f(\langle s', s, g^{-1} \rangle, u) = -t.$$

Reversibility, a property of certain processes, must not be mistaken for the invariance of a basic law statement under time reversal, i.e. under the inversion of the sign of t (or, equivalently, under the T operation). Call $L(t)$ a law statement involving a time coordinate t. If $L(-t) = L(t)$, the law is said to be *T-invariant*, or invariant under time reversal. The basic electromagnetic laws (Maxwell's) are T-invariant, whereas the fundamental law of heat transfer (Fourier's) is not. The time inverse of a process described by a T-invariant law statement is just the reverse of the original (reversible) process – e.g. a motion with inverted velocities and spins. No "inversion of time" is involved even if the process can be described by exchanging t for $-t$ in the derived equations. The reverse or transpose of a reversible process goes forward in time just as much as the original process does. There is no *arrow of time* except as a misleading and overrated metaphor. The only "arrows" can be found in certain processes – namely the irreversible ones such as radioactive disintegration or ageing. Therefore the search for the origin of the arrow of time – in which many distinguished physicists have engaged – is as futile as the search for the philosopher's stone.

If a law statement fails to be T-invariant, i.e. if $L(-t) \neq L(t)$, then it concerns irreversible processes – provided the law is basic rather than derived. (E.g., whereas the equation of motion of a vibrating string is T-invariant, its elementary solution $u = a \sin(kx - \omega t)$, a derived law, is not.) The converse is false: some irreversible processes can be described with the help of T-invariant laws together with certain subsidiary conditions. For example the basic formulas of classical electromagnetism are T-invariant but, when conjoined with Sommerfeld's condition of outward radiation (exclusion of incoming waves), they describe the irreversible propagation of a retarded outgoing wave. In general a process or history is described jointly by a set of law statements and a set of constraints, initial conditions, boundary conditions, and possibly other subsidiary hypotheses representing particular circumstances of both the system and its environment. Thus radioactive processes are described by elementary quantum mechanics, which is T-invariant, together with the hypothesis that a nuclear potential barrier is semipermeable, in the sense that some nuclear particles can tunnel through it even if their energy is smaller than the height of the barrier. In sum the T-invariance of a bunch of law statements provides no indication as to the reversibility or otherwise of the process concerned: only the whole theory can supply such an indication (cf. Bunge, 1970b, 1972).

5.3. *Antecedence* (*"Causality"*) *Principle*

Contemporary physics asserts explicitly, in various forms and under the misnomer *causality condition* (or *relation*), an overriding metaphysical principle, namely clause (vii) of Postulate 6.1. This principle states that causes precede their effects or, equivalently, that the present is determined (whether uniquely or stochastically) by the past, never by the future – this relation being moreover frame invariant unlike the mere succession of states. We proceed to state this principle in terms of the property concept and shall distinguish two cases, the stochastic and the non-stochastic ones (Bunge, 1968c).

Let P_1 and P_2 be two time dependent properties of things of a certain kind K. As usual we assume that the properties are represented by functions and, more particularly, that both $F_1 \hat{=} P_1$ and $F_2 \hat{=} P_2$ are defined on $K^n \times \{f\} \times T$, where f is a reference frame, T the range of the metric duration function attached to f, and $n \geqslant 1$ a natural number. (Actually there may be more than one kind of thing, so K may have to be interpreted as a thing genus or else K may have to be split into a cartesian product.) As far as pure mathematics is concerned there are three possibilities as to the relations between the two properties:

(*a*) P_1 and P_2 are mutually independent;

(*b*) one of them depends upon the other – e.g. P_1 determines P_2;

(*c*) both P_1 and P_2 depend upon a third property, so that they are indirectly related to one another.

Here we are interested in case (*b*), i.e. that of one-sided dependence. Again, as far as pure mathematics is concerned, there are no limitations on the form of dependence. But it so happens that actually the dependence forms are restricted by a general principle according to which the value of F_2 at any given time is the outcome of the P_1-history of the thing prior to that time. In other words the dependence of any one property on another is *retarded* or *nonanticipatory*: there is no action by the future – hence no precognition, i.e. no information flow coming from future events. This fundamental ontological hypothesis can be formulated as follows:

POSTULATE 6.12 Let P_1 and P_2 be properties of a system composed of things of kind K and assume that P_1 determines P_2.

(i) If P_1 and P_2 are representable by ordinary functions F_1 and F_2 respectively, both defined on the domain $K^n \times \{f\} \times T$, then the instantaneous value $F_2(x, y, \ldots, z, f, t)$ of F_2 for $x, y, \ldots, z \in K$, relative to the frame f, is a functional G of the P_1-history of the thing up to the given instant:

$$F_2(x, y, \ldots, z, f, t) = \underset{u=-\infty}{\overset{u=t}{G}} [F_1(x, y, \ldots, z, f, u)];$$

(ii) if P_2 is representable by a random variable F_2 (whether or not F_1 is one), then the probability that, at time t, its values lie in a preassigned interval $[a, b]$ is a functional of the entire P_1-history of the thing up to the given instant:

$$Pr[F_2(x, y, \ldots, z, f, t) \in [a, b]] = \underset{u=-\infty}{\overset{u=t}{G}} [F_1(x, y, \ldots, z, f, u)]$$

Simpler forms of the preceding axiom are obtained by forgetting that all substantial properties are defined for things and are often relative to a reference frame. I.e., by calling the property-representing functions x and y respectively, the antecedence principle simplifies to

$$\left.\begin{matrix} y(t) \\ Pr[y(t) \in [a, b]] \end{matrix}\right\} = \underset{u=-\infty}{\overset{u=t}{G}} [x(u)].$$

Actually for every thing but the whole universe, the value of G will be nil between $-\infty$ and the instant t_0 of its coming into being.

About the simplest and most common functional dependence is this:

$$\left.\begin{matrix} F_2(x, y, \ldots, z, f, t) \\ Pr[F_2(x, y, \ldots, z, f, t) \in [a, b]] \end{matrix}\right\} =$$

$$= \int_{-\infty}^{t} K(t, u) \cdot F_1(x, y, \ldots, z, f, u)\, du.$$

The particular case of instantaneous dependence of P_2 upon P_1, or instantaneous action, is obtained by setting $K(t, u) = \delta(t-u)$, where δ is the Dirac delta.

The antecedence principle has interesting consequences. One of them is

THEOREM 6.6 There is no return to the past.

Proof If there were (timelike) loops in spacetime it would be possible to modify the past states of things from the future – contrary to Postulate 6.12.

In other words, time machines are impossible. If they were possible then we could land in the past and mend it so as to obtain a better present. But then we would get two different presents. One seems enough.

Note that "time travel" to the past has nothing to do with reversibility. Even a strictly reversible universe would be subject to the restriction that the past cannot be undone. So much so that the antecedence or "causality" principle is part and parcel of quantum field theory and other stochastic theories dealing with reversible processes. Nor does Theorem 6.6 have any bearing on the frame dependence of duration as exemplified by the travelling twin "paradox". The latter consists in using the expansion of duration with motion in order to land in the future of a subject who has not taken part in the trip. This consequence of special relativity is counterintuitive to anyone brought up in the tradition of absolute time; but of course it is natural in the context of the theory and moreover it has been experimentally confirmed with great accuracy. But it is of dubious ontological value except in liberating us from the absolute theories of time.

Finally a few remarks. First, the antecedence principle is sometimes stated in this form: "Whereas the future can be influenced the past cannot". Actually neither past nor future can be influenced at any given time because they do not exist at that time (relative to a given frame). 'Influencing the future' is short and misleading for "Altering the odds of the present possibilities". There is no direct action of past on future: rather than jumping across a stretch of time we must let things unfold themselves relative to each other – or, as ordinary language puts it, in the course of time.

A second remark is this: the name 'causality condition' for Postulate 6.12 is mistaken on several counts (Bunge, 1959, Ch. 3, Sec. 3.2.) For one thing the postulate applies not just to input-output relations, which sometimes are genuinely causal, but also to certain pairs of functions mirroring different traits of one and the same thing – hence to situations in which there are no input and output terminals.

Third observation: although there is no experience of anticipation – but only hallucinations of such – some theories do allow for it at their own risk. The only great theory which is partially unfaithful to the antecedence principle is classical electrodynamics, according to which an electron is accelerated by an incoming wave before the latter actually strikes it. But this is usually regarded as one of the various unsatisfactory features of the theory, not just because it is unsupported by observation but also because it contradicts the general principle of antecedence. (See Bunge, 1967b, p. 167.)

Fourth: goal-directed behavior is only in seeming contradiction to antecedence. Indeed if an organism seems to be guided by its own future states it is only because the organism is equipped to behave in such a way that it will exploit its present circumstances so as to maximize its chances of survival or increase its welfare. Even human teleology is present-guided not future-guided: an anticipation (e.g. programming) of future events is nothing but the formation, at the present time, of a mental picture of certain possible future events, in particular those we wish to bring about.

5.4. *Action by Contact*

As far as logic is concerned a thing may act upon a separate (but of course connectible) thing either directly (*action at a distance*) or through the intermediary of a third thing (*nearby action*). And in either case such an influence could be exerted instantaneously or it could take some time. (See Figure 6.12.)

Action at a distance was assumed in all of the classical atomic theories, while nearby action has been a component of the plenistic cosmologies of Aristotle, Descartes and their followers. Until the

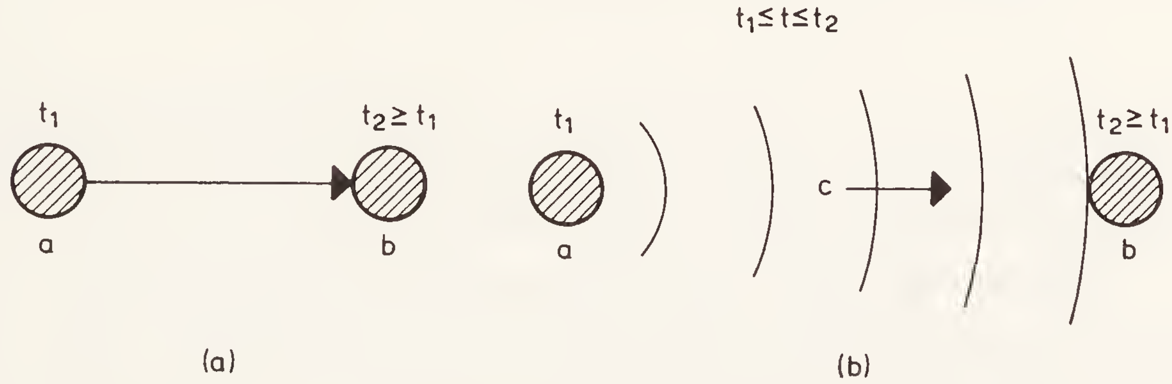


Fig. 6.12. (*a*) Thing *a* acts directly upon thing *b* (instantaneously or with some delay). (*b*) Thing *a* acts upon thing *b* through thing *c* (instantaneously or with some delay).

confirmation of the first successful field theory there was no compelling reason for preferring one of the alternatives to the other. Ever since then the hypothesis of nearby action has prevailed in every chapter of physics where interactions among different things are studied. True, corpuscularian metaphysics is still going strong among operationists who insist that we never measure anything but attributes of perceptible bodies such as the position of the hands of a clock. Here and there operationism raises its head and indicts the very concept of field, but only temporarily: it goes down every time because action-at-a-distance theories are definitely poorer than nearby action theories – in particular they cannot explain away processes which, like the propagation of an electromagnetic wave, are irreducibly plenistic. We bet then on nearby action. To formulate the principle we make use of the notion of place or bulk of a thing, introduced by Definition 6.6 in Sec. 2.4. The *principle of nearby action* reads then:

POSTULATE 6.13 If two things are at different places and one of them acts upon the other, then there is a third thing interposed between them. I.e.,

$$x, y \in \Theta \;\&\; \beta(x) \cap \beta(y) = \varnothing \;\&\; x \rhd y \Rightarrow$$
$$(\exists z)(z \in \Theta \;\&\; z \neq x \;\&\; z \neq y \;\&\; x|z|y).$$

A particularly appealing if specific version of this principle is the following. Let P_1 and P_2 be two substantial properties represented by functions F_1 and F_2 respectively defined on the space manifold M^3. Then for any point $x \in M^3$ there is a real valued function L on $M^3 \times M^3$ such that

$$F_2(x) = \int_{-\infty}^{\infty} dy\, L(x, y) F_1(y).$$

The principle of nearby action does not preclude the possibility that an action initiated in a thing remains confined to it. Such a possibility is denied by the strong version of the principle, which can be formulated thus: "Whatever happens at any given place is determined by what happens in its immediate vicinity and in turn influences every other thing". According to this principle no two things in the universe can be independent. We do not adopt it because there are cases of effective insulation and of spontaneous (uncaused) activity (Bunge, 1959).

5.5. *Spatiotemporal Contiguity*

The principles of antecedence and nearby action are logically independent: as suggested in the previous section one could assert the one while denying the other. But, since we have accepted both, we may as well merge them into a single principle, the more so since relativistic physics has taught us that spatial relations are enmeshed with temporal relations. A possible merger is the so called principle of *local causality*, which we prefer to call the *principle of spacetime contiguity*:

POSTULATE 6.14 Every action of one thing upon another satisfies both the antecedence and the nearby action conditions.

A possible but restricted formulation of this condition is as follows. If P_1 and P_2 are substantial properties of some thing (e.g. a field) representable by functions F_1 and F_2 respectively, defined on the spacetime manifold M^4, and if P_2 depends upon P_1, then relative to any frame f and for any quadruple $\langle x, t\rangle \in M^4$, there is a real-valued function N such that

$$F_2(x, t) = \int_{-\infty}^{t} dv \int_{-\infty}^{\infty} du\, N(x, u; t, v) F_1(u, v),$$

where

$$\langle u, v\rangle \in M_f^4$$

and $|x-u| \leq c|t-v|$, c being a positive real number. (To comply with special relativity c must be smaller than or equal to the velocity of light in vacuo. General relativity introduces certain modifications in the above formula, mainly the insertion of $g^{1/2}$ in the integrand, where g is the determinant of the metric tensor, and a slight liberation of the constraint on space and duration intervals, as c may now be a function defined on M^4, though one whose values will always be near the velocity of light in vacuo in the absence of gravitation.)

The previous principles, of nearby action and antecedence, are retrieved as special cases upon setting $N(x, u; t, v) = \delta(t-v) \cdot L(x, u)$ and $N(x, u; t, v) = \delta(x-u) \cdot K(t, v)$ respectively, where L and K are space dependent and time dependent functions respectively.

To realize that Postulate 6.14 underlies the whole of science suffice it to note that its denial would involve the following consequences. First, because the principle asserts the mutual independence of nonconnectible things, if it were false no thing could remain quasi-isolated from the rest of the universe even for a short while. Hence no thing could be known even approximately, and so the very concept of an individual thing would be idle. Far from being a system, the universe would be a block. Second, both the remote past and the remote future could pounce upon the present without notice and render it chaotic. This again would render knowledge utterly impossible and confusion would be even more dense and widespread than it is. Since we do succeed in knowing a thing or two we must infer that we are not at the mercy of things far removed in space and time. In particular, we may rest assured that events occurring in a distant space-like separated region of spacetime do not affect us.

True, it is occasionally claimed that Postulate 6.14 is at variance with quantum mechanics. But if this were true then the quantum theory would be incompatible with the special theory of relativity, which in turn embodies Postulate 6.14. And this conclusion is contradicted by the very existence of successful quantum relativistic theories such as Dirac's theory of the electron and quantum electrodynamics. Hence the results purporting to invalidate the so-called local causality – such as EPR inseparability and Bell's theorem – will have to be reinterpreted. In other words we do not have to give up ontological principles that are well entrenched in science just because they seem to be contradicted by an occasional and isolable fragment of scientific research. On the contrary, such strays have got to be reexamined in the light of scientific ontology. Thus when classical electrodynamics led to the conclusion that there are preaccelerations, hence events violating the antecedence principle (Postulate 6.12), most physicists doubted the physics rather than the metaphysics and made efforts to correct the former instead of the latter. (Recall Sec. 5.3.)

5.6. *The Causal Relation*

Finally let us elucidate the notion of causation and state the causal principle. We construe causation as a relation between events in different things – which of course may be parts of a bulkier thing. Furthermore we take it that causation subsumes antecedence and that, far from merely preceding the effect, the cause produces it (Bunge 1959). In sum, we make

DEFINITION 6.23 Let $e \in E(x)$ be an event in a thing x at time t and $e' \in E(x')$ another event in a thing $x' \neq x$ at time t', where the events as well as the times are taken relative to the same reference frame. Further, call $A(x, x')$ the total action or effect of x upon x' (Definition 5.31). Then we say that e is a *cause* of e' iff

(i) $t \leqslant t'$;

(ii) $e' \in A(x, x') \subseteq E(x')$.

A more precise condition than clause (ii) would be this: There is an energy transfer from x to x' during the interval $[t, t']$, and the quantity of energy transferred suffices for the occurrence of e'. But we do not need the extra precision.

The notion of causation allows us to formulate the strict *principle of causality*, namely as follows: "Every event is caused by some other event." More precisely: "Let x be a thing with event space $E(x)$. Then for every $e \in E(x)$ there is another thing $x' \neq x$, with event space $E(x')$ relative to the same reference frame, such that $e' \in E(x')$ causes e." But we shall not espouse this principle for, if we did, most of contemporary science would have to go (Bunge, 1959).

So much for the general principles involving the notions of space or time. Let us now approach matters of existence.

6. MATTERS OF EXISTENCE

6.1. *Existence in Space and Time*

It is usually assumed that, unlike ideas, physical things "exist in space and time". (For Aristotle only changing things exist in time: *Physics*, Bk. IV, Ch. 12. And for Kant ideas exist in time but not in space.) Moreover "existence in space and time" is sometimes taken to define the very concept of a physical object in contrast to that of a nonphysical object. In turn, what is meant by 'x exists in space and time' is that x occupies a region of space and endures throughout a time interval.

The statements we have just mentioned presuppose the autonomous existence of the spatiotemporal framework, which in turn would be a non-physical object. This is of course the absolute doctrine examined and rejected in Sec. 1.1. We do not accept the container view because our ontology does not assume any nonphysical entity – except of course as a fiction. In our view things do not float in a given spatiotemporal framework but hold spatiotemporal relations among each other – just as persons do not swim in a community but constitute it by holding connections with one another. Things "come" with their own (changing) spatiotemporal relations, and the latter are just relations (but not bonds) among things and their changes.

Consequently it makes as little sense, in our ontology, to say that a thing *exists in* spacetime, as to say that spacetime *exists in* a thing. The expressions 'x exists at a place y (or in a region y) relative to a frame z' and 'x exists at time t (or throughout time interval t) relative to a frame z' are elliptical. The location of a thing is always given by reference to another thing and the timing of a state or of an event is always referred to some other state or event – and in turn states are states of things, and events are changes of state of things. *No things, no nothing*, as the robust street corner philosopher put it. We can use those elliptical locutions because they save time, provided we remember what their purport is at the time of discussing foundations matters. In ontology and in the foundations of physics we have hardly any right to assume that spacetime preexists things, and that the spatiotemporal location of a thing or an event is in the nature of the part-whole relation, the part being the factual item and the whole spacetime. Only the bulk of a thing can be included in space, because that bulk is a set.

Finally, what about the *plenum* hypothesis held by Aristotle and his followers? (Cf. Lovejoy, 1936.) This hypothesis, that space is full of things, is favored by contemporary physics, in particular gravitation theory and quantum electrodynamics. In other words, there is no such thing as empty space or a perfect vacuum. (Particle void $\neq$ void.) We may then adopt this principle:

POSTULATE 6.15 There are things interposed between any two things located at different places. I.e.,

$$x, y \in \Theta \;\&\; \beta(x) \cap \beta(y) = \varnothing \Rightarrow$$
$$(\exists z)(z \in \Theta \;\&\; z \neq x \;\&\; z \neq y \;\&\; x|z|y).$$

Since this assumption entails Postulate 6.13, we could have gotten the latter as a theorem.

6.2. *Existence of Space and Time*

It is tempting to doubt the existence of space and time. After all, they are not things – not even imperceptible things – and they have no causal power. (It is sometimes said that, according to general relativity, spacetime guides the motion of things: that there are tracks in spacetime. But this is a mistake: what guides the motion of things is the gravitational field, which the theory represents as a departure of spacetime from flatness.)

As a matter of fact a number of philosophers and mystics have doubted the reality of space and many more have maintained that time is unreal. While the topoclasts have usually argued from the imperceptibility of space, the chronoclasts have often argued from the alleged immutability of a Reality behind appearances. In particular Kant held space and time to be phenomenal not real – which is the only reasonable view if the container doctrine is kept, not however on a relational doctrine of space and time (Lotze, 1887, Bk. II, Ch. I). And Russell held at one time that both space and time are fictitious because the *concepts* of space and time should be constructible out of "real entities", namely perceptual items. We cannot accept these phenomenalist views if only because they ignore the basic distinction made in contemporary physiological psychology between physical spacetime on the one hand and the various perceptual (e.g. visual and tactual) spaces and the perception of duration on the other.

In our view neither the topoclasts nor the chronoclasts are right. Although space and time are not things and have no being independent of things and their changes, they are no less real than any other substantial properties of real things. Indeed recall that a relation was called *real* if and only if it holds among real things – which is the case with spatiotemporal relations. (Cf. Definition 2.17 in Ch. 2, Sec. 5.1.) The *esse* of spatiotemporal relations consists not in their *percipi* but in their obtaining between real things and real events.

The role of space and time in science is not always well understood. Thus just because spatiotemporal concepts do not occur in the systematics of elementary particles, or because they play a subdued role in the black box theories of these entities – namely the scattering matrix and the dispersion relation theories – it has occasionally been claimed that they play no role at all (Chew, 1963). However, the fact that a dynamical theory of fundamental "particles" is not around the corner does not entail that we can dispense with space and time concepts. This is a case of sour grapes. It is no less glaring a mistake than that of assigning spacetime causal efficacy – which Aristotle did in his *Physics*. The locution 'x was knocked down because x was at place y' is just short for 'x was knocked down by some other thing z that happened to be at place y'. Likewise the phrase 'Time brought about the decline of x' is an ellipsis for 'Certain events brought about the decline of x'. Space and time are as harmless as they are unavoidable.

Another point that has not been well understood on occasion is the nature of coordinates. Because the general covariance requirement shows coordinate systems and coordinate values to be conventional or a priori, it is sometimes held that spacetime itself is not physically objective (e.g. Einstein, 1916). That coordinate values just label points in the four-manifold M^4 is true and even trivially so, as we saw in Sec. 4.2. But this does not entail that M^4 fails to represent objective relations among physical objects, as shown by the absolute character of certain invariants such as spatiotemporal distances (the values of the numerical partners of the qualitative spatiotemporal separation introduced by Definition 6.18). In fact they do not change under the substitution of coordinate systems; in particular they do not depend on the observer's point of view nor even upon his location and motion. Moreover the coefficients of the metric, though not covariant themselves (since they are the components of a tensor), are no less than the gravitational potentials and, as such, they occur centrally in the field equations of gravitation: they represent a feature of the thing called 'gravitational field'. If the metric were conventional (as Grünbaum (1973) has claimed), the laws of gravitation should be conventional as well. And if this were so we should be smart enough always to choose a metric that will prevent any nasty fall and even keep our enemies at bay.

In conclusion, space and time have no autonomous existence, hence no causal power either. But then no relation exists on its own, i.e. apart from its relata – as we know from Ch. 2. Although spatiotemporal relations are not entities, nor even couplings or connections on the same footing with mechanical or electrical or social bonds, they are real relations in the sense that they hold among real things. Spacetime is no more nor less than *the fundamental structure or basic framework of the world.*

In sum: mathematical spaces are multiple and unreal – neither objective nor subjective; physical space(time) is one, real, and objective (or subject-independent); and phenomenal (or perceptual) spaces and times are multiple, real, and subjective (or subject-dependent).

7. CONCLUDING REMARKS

According to Cartesian mechanism all facts can and should be explained in terms of space and time only. This program, though wonderfully fruitful, has never been fully implemented in physics, let alone in other fields of research. Indeed some of the basic properties of matter, such as mass, charge, spin, and parity, are not explained in terms of spatiotemporal concepts. And even features that are dependent on spatiotemporal location, such as field strengths and cultural production, are not spatiotemporal themselves. In sum, things are not warts of spacetime. And spacetime, not being independent of things, is not illusory either. Reality is the aggregation of things holding spatiotemporal relations. Now, relations do not preexist their relata but hold among them. In particular spatiotemporal relations, such as "between" and "before", can be understood only in terms of the things, states, or events between which they hold. Hence we must try and understand spacetime in terms of changing things: this is what the present chapter is all about.

Once the spatiotemporal framework has been anchored to things we can invert the procedure and study things *in* spacetime – but this second, more detailed look, is taken by science not philosophy. (Likewise once we have formed the general concept of an event, belonging to ontology, we can start asking questions about events of special kinds.) In summary we have the situation depicted in Figure 6.13.

How do the various ontological theories of space and time fare in

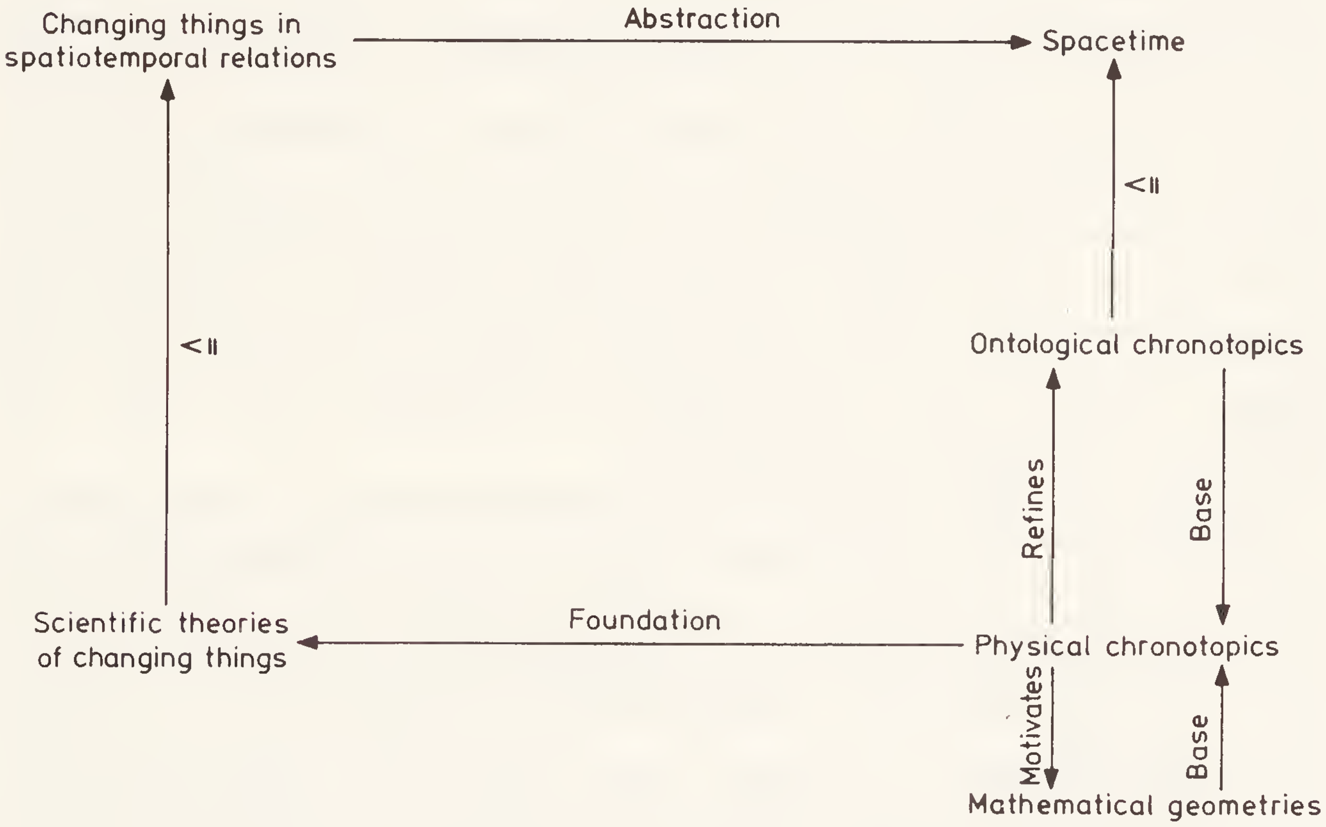

Changing things in spatiotemporal relations
Abstraction
Spacetime
<||
<||
Ontological chronotopics
Refines
Base
Scientific theories of changing things
Foundation
Physical chronotopics
Motivates
Base
Mathematical geometries

Fig. 6.13. Relations among items discussed in this chapter. The real thing is in the upper left corner: the rest are constructs.

contemporary science, i.e. what support do they get from the latter? This is not easy to say because any physical theory containing spatiotemporal concepts can presumably be married either to an absolute or to a relational theory of spacetime. Such an ambiguity, clear as it is on analysis, is obscured by history. Thus just because Newton assumed that space and time are self-existent, and in particular independent of matter in motion, most philosophers believe that Newtonian mechanics and gravitation theory are soaked in the doctrine of the absolute character of space and time. This is mistaken: it is possible to reformulate classical mechanics assuming a relational (though not relativistic) theory of space and time (Noll, 1967). In other words, the equations of motion and the constitutive equations of classical mechanics are independent of the way space and time are understood. The same holds for all other non-relativistic theories, in particular quantum mechanics.

Even relativistic theories are rather insensitive to the absolute-relational issue. In this case too either one may attempt to build spacetime out of changing things, or one may take it for granted and build thereupon. One may even get carried away by the second procedure and conclude that, since the gravitational field equations possess solutions even in the absence of matter and radiation, general relativity actually uses an absolute concept of spacetime. But this would be just as mistaken as claiming that Newtonian mechanics is indissolubly married to an absolute doctrine of space and time. The solutions for the hollow universe are physically meaningless because, by definition, physics is concerned with physical objects. In sum, the fact that physics is compatible with a given ontological theory of space and time does not endorse the latter but only renders it epistemically possible. In particular, the compatibility of our own chronotopics with the science of the day, though necessary and encouraging, is insufficient. We must look elsewhere for additional confirmation.

We cannot verify conclusively our relational theory of spacetime – nor indeed any other chronotopics. However we can dispose of its rival, namely the absolute theory, in any of its possible versions. We can do it not with the help of crucial experiments but by a semantic analysis. Indeed, if one admits the methodological principle that a physical theory, by definition of 'physical', must contain only physically meaningful concepts, and if one accepts the definition of a physically meaningful concept as one that refers to physical items (cf. Vol. 1, Ch. 2 of this treatise), then the concepts of absolute or autonomous space and time must be discarded for not being physical. By this criterion only the class of relational theories of space and time remain. And the only criteria for choosing among the latter are (*a*) mathematical cogency, (*b*) closeness of fit with contemporary physics, and (*c*) clarifying power. We leave the reader the task of deciding whether our own theory of spacetime meets these criteria.

So much for extension and duration. We have now all we need to handle systems in general and in particular. But this will be a task for the 4th volume of this work, namely *A World of Systems*.

BIBLIOGRAPHY

Agassi, J. (1964). The nature of scientific problems and their roots in metaphysics. In M. Bunge, Ed., *The Critical Approach*, pp. 189–211. New York: Free Press.

Agassi, J. (1975). *Science in Flux*. Dordrecht & Boston: Reidel.

Alexander, H. G., Ed. (1956). *The Leibniz–Clarke Correspondence*. Manchester: University Press.

Alexander, S. (1920). *Space, Time and Deity*, 2 vols. New York: Humanities Press.

Allen, J. A. (1870–71). On the mammals and winter birds of East Florida, etc. *Bull. Museum Compar. Zoölogy* (Harvard) **2**: 161–450.

Aquinas, T. (1259–64). *Summa contra gentiles*. Transl. A. C. Pegis. Garden City, N.Y.: Hanover House 1955.

Aristotle. *Metaphysics*. In *Works*, J. A. Smith and W. D. Ross, Eds., vol. VIII. Oxford: University Press 1928.

Aristotle. *Physics*. In *Works*, J. A. Smith and W. D. Ross, Eds., vol. II. Oxford: Clarendon Press 1930.

Ashby, W. R. (1956). *An Introduction to Cybernetics*. New York: McGraw-Hill.

Basri, S. A. (1966). *A Deductive Theory of Space and Time*. Amsterdam: North-Holland.

Bergmann, G. (1960). The philosophical significance of modal logic. *Mind* (N.S.) **69**: 466–485.

Bergmann, H. (1929). *Der Kampf um das Kausalgesetz in der jüngsten Physik*. Braunschweig: Vieweg.

Bergson, H. (1903). Introduction à la métaphysique. *Revue de métaphysique et de morale* **11**: 1–36.

Bernays, P. (1937). A system of axiomatic set theory. Part I. *Journal of Symbolic Logic* **2**: 65–77.

Blalock, H. M. (1961). Evaluating the relative importance of variables. *Am. Sociol. Rev.* **26**: 866–874.

Bohm, D. (1970). Further remarks on order. In C. H. Waddington, Ed., *Towards a Theoretical Biology*. Chicago: University of Chicago Press.

Bolzano, B. (1821). *Begriffe*. In E. Winter, *Die historische Bedeutung der Frühbegriffe B. Bolzanos*. Berlin: Akademie-Verlag 1964.

Bolzano, B. (1837). *Wissenschaftslehre*, 4 vols. Repr. Leipzig: Meiner 1929, and Aalen: Scientia Verlag, 1970.

Bolzano, B. (1849). *Was ist Philosophie*? Amsterdam: Rodopi 1969.

Bolzano, B. (1851). *Paradoxes of the Infinite*. Transl. by D. A. Steele. London: Routledge & Kegan Paul 1950.

Booij, H. L., and H. P. Wolvekamp (1944). *Catenary Processes, Master Reactions and Limiting Factors*. Bibliotheca Biotheoretica, Vol. I, Pars 4. Leiden: E. J. Brill.

Borel, E. (1948). *Le hasard*, 2nd ed. Paris: Presses Universitaires de France.

Bradley, F. H. (1893). *Appearance and Reality*. London: Oxford University Press 1969.

Braudel, F. (1969). *Écrits sur l'histoire.* Paris: Flammarion.

Brody, T. A. (1974). Probability: A new look at old ideas. Preprint, Instituto de Física, Universidad Nacional Autónoma de México.

Bruck, R. H. (1958). *A Survey of Binary Systems.* Berlin, Göttingen & Heidelberg: Springer-Verlag.

Buchdahl, G. (1969). *Metaphysics and the Philosophy of Science.* Cambridge, Mass.: MIT Press.

Bunge, M. (1959). *Causality.* Cambridge, Mass.: Harvard University Press.

Bunge, M. (1961). The weight of simplicity in the construction and assaying of scientific theories. *Phil. Sci.* **28**: 129–149.

Bunge, M. (1962). Cosmology and magic. *The Monist* **47**: 116–141.

Bunge, M. (1964). Phenomenological theories. In M. Bunge, Ed., *The Critical Approach.* New York: The Free Press.

Bunge, M. (1966). On null individuals. *J. Phil.* **63**: 776–778.

Bunge, M. (1967a). *Scientific Research,* 2 vols. New York: Springer-Verlag.

Bunge, M. (1967b). *Foundations of Physics.* New York: Springer-Verlag.

Bunge, M. (1968a). Physical time: the objective and relational theory. *Phil. Sci.* **35**: 355–388.

Bunge, M. (1968b). Physique et métaphysique du temps. *Proc. XIVth Intern. Cong. Philosophy* **II**: 623–629. Wien: Herder.

Bunge, M. (1968c). Conjunction, succession, determination, and causation. *Intern. J. Theoret. Phys.* **1**: 299–315.

Bunge, M. (1968d). Problems and games in the current philosophy of natural science. *Proc. XIVth Intern. Cong. Philosophy* **I**: 566–574. Wien: Herder.

Bunge, M. (1969). What are physical theories about? *Amer. Phil. Quart. Monographs* **3**: 61–99.

Bunge, M. (1970a). Virtual processes and virtual particles: real or fictitious? *Intern. J. Theoret. Phys.* **3**: 507–508.

Bunge, M. (1970b). The arrow of time. *Intern. J. Theoret. Phys.* **3**: 77–78.

Bunge, M. (1971a). Is scientific metaphysics possible? *J. Philos.* **68**: 507–520.

Bunge, M. (1971b). A mathematical theory of the dimensions and units of physical quantities. In M. Bunge, Ed., *Problems in the Foundations of Physics,* pp. 1–16. New York: Springer-Verlag.

Bunge, M. (1972). Time symmetry, time reversal, and irreversibility. In J. T. Fraser *et al.,* Eds., *The Study of Time,* pp. 122–130. New York: Springer-Verlag.

Bunge, M. (1973a). *Method, Model and Matter.* Dordrecht & Boston: Reidel.

Bunge, M. (1973b). *Philosophy of Physics.* Dordrecht: Reidel.

Bunge, M. (1973c). Review of P. Suppes' *Probabilistic Theory of Causality. Brit. J. Phil. Sci.* **24**: 409–410.

Bunge, M. (1974a). *Sense and Reference.* Vol. 1 of *Treatise on Basic Philosophy.* Dordrecht & Boston: Reidel.

Bunge, M. (1974b). *Interpretation and Truth.* Vol. 2 of *Treatise on Basic Philosophy.* Dordrecht & Boston: Reidel.

Bunge, M. (1974c). The relations of logic and semantics to ontology. *Journal of Philosophical Logic* **3**: 195–209.

Bunge, M. (1974d). Les présupposés et les produits métaphysiques de la science et de la technique contemporaines. *Dialogue* **13**: 443–453.

Bunge, M. (1974e). The concept of social structure. In W. Leinfellner and E. Köhler, Eds., *Developments in the Methodology of Social Science*, pp. 175–215. Dordrecht & Boston: Reidel.

Bunge, M. (1974f). Metaphysics and science. *General Systems* **19**: 15–18.

Bunge, M. (1974g). Things. *Intern. J. General Systems* **1**: 229–236.

Bunge, M. (1975). A critical examination of dialectics. In Ch. Perelman, Ed., *Dialectics/Dialectique*, pp. 63–77. The Hague: Nijhoff.

Bunge, M. (1976a). Possibility and probability. In W. Harper and C. Hooker, Eds., *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, Vol. III, 17–33. Dordrecht and Boston: D. Reidel Publ. Co.

Bunge, M. (1977a). The GST challenge to the classical philosophies of science. *Intern. J. General Systems*, Vol. 4, No. 1.

Bunge, M. (1977b). States and events. In W. E. Hartnett, Ed., 1977.

Bunge, M. (1977c). The interpretation of Heisenberg's inequalities. In H. Pfeiffer, Ed., *Denken und Umdenken: Zu Werk und Wirkung von Werner Heisenberg*. München–Zürich: R. Piper & Co.

Bunge, M., and A. García Máynez (1977). A relational theory of physical space. *Intern. J. Theoret. Phys.*, Vol. 16, forthcoming.

Bunge, M., and A. A. L. Sangalli (1977). A theory of properties and kinds. *Intern. J. General Systems*: **3**: 183–190.

Burtt, E. A. (1932). *The Metaphysical Foundations of Physical Science*, rev. ed. London: Routledge & Kegan Paul.

Carnap, R. (1928). *Der logische Aufbau der Welt.* Transl. by R. A. George: *The Logical Structure of the World.* Berkeley & Los Angeles: University of California Press 1967.

Carnap, R. (1936–37). Testability and meaning. *Phil. Sci.* **3**: 419–471; **4**: 1–40.

Carnap, R. (1947). *Meaning and Necessity.* Chicago: University of Chicago Press.

Castañeda, H.-N. (1974). Thinking and the structure of the world. *Philosophia* **4**: 3–40.

Chaitin, G. J. (1974). Information-theoretic limitations of formal systems. *J. Assoc. for Computing Machinery* **21**: 403–424.

Chew, G. (1963). The dubious role of the space-time continuum in microscopic physics. *Sci. Progr.* **51**: 529–539.

Chew, G. (1966). *The Analytic S Matrix.* New York: Benjamin.

Chisholm, R. M. (1960). *Realism and the Background of Phenomenology.* Glencoe, Ill.: The Free Press.

Chwistek, L. (1949). *The Limits of Science.* London: Routledge & Kegan Paul.

Cocchiarella, N. B. (1972). Properties as individuals in formal ontology. *Noûs* **6**: 165–187.

Cocchiarella, N. B. (1974). Logical atomism and modal logic. *Philosophia* **4**: 41–66.

Collingwood, R. G. (1940). *An Essay on Metaphysics.* Oxford: Clarendon Press.

Costa de Beauregard, O. (1963). *Le second principe de la science du temps.* Paris: Ed. du Seuil.

Cournot, A.-A. (1851). *Essai sur les fondements de nos connaissances.* Eng. transl. with Introd. by M. H. Moore: *An Essay on the Foundations of our Knowledge.* New York: Liberal Arts Press 1956.

Cresswell, M. J. (1973). Physical theories and possible worlds. *Logique et analyse* (N.S.) **16**: 495–511.

Dantzig, D. van (1935). The fundamental equations of electromagnetism, independent of metric geometry. *Cambridge Phil. Soc. Proc.* **30**: 421–427.

Dehnen, H. (1970). The frame of reference in relativistic theories. *Intern. J. Theoret. Phys.* **3**: 509–511.

Dewitt, B. S. (1970). Quantum mechanics and reality. *Physics Today* **23**, No. 9, pp. 30–35.

DeWitt, N. W. (1954). *Epicurus and his Philosophy*. Minneapolis: University of Minnesota Press.

Ducasse, C. J. (1968). *Truth, Knowledge and Causation*. New York: Humanities Press.

Einstein, A. (1916). Grundlagen der allgemeinen Relativitätstheorie. *Ann. Phys.* **49**: 769–822.

Engels, F. (1878). *Anti-Dühring*. In K. Marx and F. Engels, *Werke*, vol. 20. Berlin: Dietz 1973.

Everett, H., III (1957). 'Relative state' formulation of quantum mechanics. *Rev. Mod. Phys.* **29**: 454–463.

Falk, G. (1966). *Theoretische Physik, I*. Berlin, Heidelberg & New York: Springer-Verlag.

Feibleman, J. and J. W. Friend (1945). The structure and function of organization. *Phil. Rev.* **54**: 19–44.

Feigl, H. (1967). *The 'Mental' and the 'Physical'*. Minneapolis: University of Minnesota Press.

Feller, W. (1968). *An Introduction to Probability Theory and its Applications*, I, 3rd. ed. New York: Wiley.

Fine, A. (1964). Physical geometry and physical laws. *Phil. Sci.* **31**: 156–162.

Finkelstein, D. (1973). A process conception of nature. In J. Mehra, Ed., *The Physicist's Conception of Nature*, pp. 709–713. Dordrecht: Reidel.

Fitch, F. B. (1971). Propositions as the only realities. *Amer. Phil. Quart.* **8**: 99–103.

Fokker, A. D. (1965). *Time and Space, Weight and Inertia*. Oxford: Pergamon Press.

Fraassen, B. C. van (1970). *An Introduction to the Philosophy of Space and Time*. New York: Random House.

Fréchet, M. (1939). The diverse definitions of probability. *Erkenntnis* **8**: 7–23.

Gaal, S. A. (1964). *Point Set Topology*. New York: Academic Press.

Gell-Man, M. (1956). The interpretation of the new particles as displaced charge multiplets. *Nuovo Cimento Suppl.* (S.10) **4**: 848–866.

Giere, R. N. (1974). Objective single-case probabilities and the foundations of statistics. In P. Suppes *et al.*, Eds., *Logic, Methodology and Philosophy of Science IV*, pp. 467–483. Amsterdam: North-Holland 1974.

Gold, T., Ed. (1967). *The Nature of Time*. Ithaca, N.Y.: Cornell University Press.

Gonseth, F. (1938). *La méthode axiomatique*. Paris: Gauthier-Villars.

Goodman, N. (1951). *The Structure of Appearance*. Cambridge, Mass.: Harvard University Press.

Goodman, N. (1956). A world of individuals. In *The Problem of Universals*, pp. 15–31. Notre Dame, Ind.: University of Notre Dame Press.

Grünbaum, A. (1967). *Modern Science and Zeno's Paradoxes*. Middleton, Conn.: Wesleyan University Press.

Grünbaum, A. (1973). *Philosophical Problems of Space and Time*. Dordrecht & Boston: Reidel.

Harary, F. (1971). Demiarcs: an atomistic approach to relational systems and group dynamics. *J. Math. Sociology* **1**: 195–205.

Harré, R. (1961). *Theories and Things*. London: Sheed and Ward.

Harré, R. (1972). *The Philosophies of Science*. London and New York: Oxford University Press.

Harris, E. E. (1965). *The Foundations of Metaphysics in Science.* New York: Humanities Press.

Hart, R. G. (1971). A ramiform view of unpredictable events. *Am. J. Phys.* **39**: 1336–1346.

Hartmann, N. (1938). *Möglichkeit und Wirklichkeit.* Berlin: W. de Gruyter.

Hartnett, W., Ed. (1977). *Systems: Approaches, Theories, Applications.* Dordrecht and Boston: Reidel.

Harvey, D. (1969). *Explanation in Geography.* New York: St. Martin's Press.

Hasenjäger, G. (1966). Logik und Ontologie. *Studium generale* **19** (3): 136–140.

Hebb, D. (1949). *The Organization of Behavior.* New York: Wiley.

Hegel, G. W. F. (1812–16). *The Science of Logic.* Transl. A. V. Miller. London: Allen & Unwin 1969.

Heidegger, M. (1927). *Sein und Zeit.* Eng. transl. J. Macquarrie and E. S. Robinson: *Being and Time.* New York: Harper 1962.

Heidegger, M. (1953). *Einführung in die Metaphysik.* Tübingen: Niemeyer.

Heisenberg, W. (1969). *Der Teil und das Ganze.* München: Piper.

Helmholtz, H. von (1873). *Popular Lectures on Scientific Subjects.* London: Longmans, Green & Co.

Hilbert, D. (1899). *Grundlagen der Geometrie.* Leipzig. Eng. transl. E. J. Townsend: *Foundations of Geometry.* La Salle, Ill.: Open Court, 1902.

Hintikka, J. (1969). *Models for Modalities.* Dordrecht: Reidel.

Hirsch, M. and S. Smale (1974). *Differential Equations, Dynamical Systems, and Linear Algebra.* New York: Academic Press.

Hiż, H. (1971). On the abstractness of individuals. In M. K. Munitz, Ed., *Identity and Individuation,* pp. 251–261. New York: New York University Press.

Hooker, J. D. (1853–55). *Flora Novae Zelandae,* being v. 2 of *The Botany of the Antarctic Voyage of H.M. Discovery Ships* Erebus *and* Terror *in the Years 1839–1843,* etc. London: Reeve Brothers 1844–1860.

Hughes, G. E., and M. J. Cresswell (1968). *An Introduction to Modal Logic.* London: Methuen.

Hume, D. (1739–1740). *A Treatise of Human Nature,* L. A. Selby-Bigge, Ed. Oxford: Clarendon Press, 1888.

Huntington, E. V. (1913). A set of postulates for abstract geometry, expressed in terms of the simple relation of inclusion. *Math. Ann.* **73**: 522–559.

Jammer, M. (1954). *Concepts of Space.* Cambridge, Mass.: Harvard University Press.

Jensen, U. J. (1972). Conceptual epiphenomenalism. *The Monist* **56**: 250–275.

Kant, I. (1781). *Kritik der reinen Vernunft.* Hamburg: F. Meiner 1952.

Keynes, J. M. (1921). *A Treatise on Probability.* London: Macmillan.

Klir, G., Ed. (1972). *Trends in General Systems Theory.* New York: Wiley-Interscience.

Kneale, W. and M. (1962). *The Development of Logic.* Oxford: Clarendon Press.

Körner, S. (1970). *Categorial Frameworks.* Oxford: Basil Blackwell.

Koyré, A. (1957). Galileo and Plato. In P. P. Wiener and A. Noland, Eds., *Roots of Scientific Thought,* pp. 147–175. New York: Basic Books.

Kripke, S. (1959). A completeness theorem in modal logic. *J. Symbolic Logic* **24**: 1–15.

Kripke, S. (1971). Naming and necessity. In D. Davidson and G. Harman, Eds., *Semantics of Natural Languages.* Dordrecht: D. Reidel Publ. Co.

Kulakov, Ju. I. (1970). The geometry of spaces of constant curvature as a special case of the theory of physical structures. *Sov. Math. Dokl.* **11**: 1055–1057.

Lakatos, I. (1969). Criticism and the methodology of scientific research programs. *Proc. Aristotelian Soc.* **69**: 149–186.

Lambert, J. H. (1770). Letter to I. Kant, 13.10.1770. In H. E. Fischer, Ed., *Briechwechsel von Imm. Kant,* 3 vols. München: Georg Müller 1912.

Leibniz, G. W. von (1704). *Nouveaux Essais.* Paris: Flammarion s.d.

Leibniz, G. W. von (1956). *Philosophical Papers and Letters,* L. H. Loemker, Ed., 2 vols. Chicago: University of Chicago Press.

Lejewski, C. (1967). A single axiom for the mereological notion of proper part. *Notre Dame Journal of Formal Logic* **8**: 279–285.

Leonard, H. S. and N. Goodman (1940). The calculus of individuals and its uses. *J. Symbolic Logic* **5**: 45–55.

Lévi-Strauss, C. (1958). *Anthropologie structurale.* Paris: Plon.

Lewes, G. H. (1879). *Problems of Life and Mind.* London: Trubner.

Lewis, C. I. (1923). Facts, systems, and the unity of the world. *Journal of Philosophy* **20**: 141–151.

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation.* La Salle, Ill.: Open Court.

Lewis, C. I. and C. H. Langford (1932). *Symbolic Logic.* New York: Century.

Lewis, D. (1973). *Counterfactuals.* Oxford: Basil Blackwell.

Ljapin, E. S. (1963). *Semigroups.* Providence, R. I.: American Mathematical Society.

Locke, J. (1689). *An Essay Concerning Human Understanding.* London: Routledge s.d.

Lotze, H. (1887). *Metaphysics,* 2 vols., 2nd ed. Transl. ed. by B. Bosanquet. Oxford: Clarendon Press.

Lovejoy, A. O. (1936). *The Great Chain of Being.* Cambridge, Mass.: Harvard University Press.

Lucas, J. R. (1973). *A Treatise on Time and Space.* London: Methuen.

Luce, R. (1960). The theory of selective information and some of its behavioral applications. In R. D. Luce, Ed., *Developments in Mathematical Psychology,* pp. 5–119. Glencoe, Ill.: Free Press.

Lucretius (55 B.C.?). *De rerum natura.* Eng. transl. by W. E. Leonard: *Of the Nature of Things.* London: J. M. Dent; New York: E. P. Dutton, 1921.

Ludwig, G., Ed. (1964). *Raum und Zeit als Probleme der Naturwissenschaften.* Freiburg: Karl Alber.

Łukasiewicz, J. (1970). *Collected Works,* L. Borkowski, Ed. Amsterdam: North-Holland.

Mach, E. (1872). *Die Geschichte und die Wurzel des Satzes von der Erhaltung der Arbeit,* 2nd ed. Leipzig: Barth 1909.

Mach, E. (1883). *Die Mechanik in ihre Entwicklung historisch-kritisch dargestellt,* 4th ed. Leipzig: Brockhaus 1901.

Mach, E. (1906). *Die Analyse der Empfindungen,* 5th ed. Jena: Gustav Fischer.

Marcus, R. B. (1968). Modal logic. In R. Klibansky, Ed., *Contemporary Philosophy,* vol. I, pp. 87–101. Firenze: La Nuova Italia Editrice.

Margenau, H. (1941). Metaphysical elements in physics. *Rev. Mod. Phys.* **13**: 176–189.

Margenau, H. (1950). *The Nature of Physical Reality.* New York: McGraw-Hill Book Co.

Margenau, H. (1966). Exclusion principle and measurement theory. In P. O. Löwdin, Ed., *Quantum Theory of Atoms, Molecules, Solid State,* pp. 81–91. New York: Academic Press.

Martin, R. M. (1965). Of time and the null individual. *J. Phil.* **62**: 723–736.

Maxwell, J. C. (1892). *A Treatise on Electricity and Magnetism,* 3rd ed. Oxford: University Press.

McGilvary, E. B. (1951). Space-time, simple location, and prehension. In P. A. Schilpp, Ed., *The Philosophy of Alfred North Whitehead*, 2nd, ed. New York: Tudor Publ. Co.

Meinong, A. (1904). *The Theory of Objects*. In R. M. Chisholm (1960) pp. 76–117.

Meinong, A. (1915). *Über Möglichkeit und Wahrscheinlichkeit*. Leipzig: J. A. Barth.

Mellor, D. H. (1971). *The Matter of Chance*. Cambridge: University Press.

Melvin, M. A. (1960). Elementary particles and symmetry principles. *Rev. Mod. Phys.* **32**: 477–518.

Mendelson, E. (1963). *Introduction to Mathematical Logic*. Princeton: Van Nostrand.

Mesarović, M. D., and Y. Takahara (1975). *General Systems Theory*. New York: Academic Press.

Minkowski, H. (1908). Space and time. In H. A. Lorentz *et al.*, *The Principle of Relativity*, pp. 73–91. New York: Dover s.d.

Misner, C. W., K. S. Thorne and J. A. Wheeler (1973). *Gravitation*. San Francisco: Freeman.

Montagu, W. P. (1925). *The Ways of Knowing*. London: Allen & Unwin.

Montague, R. (1960). Logical necessity, physical necessity, ethics and quantifiers. *Inquiry* **3**: 259–269.

Montague, R. (1974). *Formal Philosophy*. Selected Papers of R. M. ed. by R. Thomason. New Haven: Yale University Press.

Munitz, M. K., Ed. (1971). *Identity and Individuation*. New York: New York University Press.

Munitz, M. K., Ed. (1973). *Logic and Ontology*. New York: New York University Press.

Naess, A. (1972). *The Pluralist and Possibilist Aspect*. Oslo: Universitetsforlaget; London: Allen & Unwin.

Neumann, J. von (1932). *Mathematische Grundlagen der Quantenmechanik*. Berlin: Julius Springer.

Newton, I. (1704). *Opticks*. In *Opera quae exstant omnia*, vol. IV. London: Nichols 1782.

Nicod, J. (1923). *La géométrie dans le monde sensible*. Paris: Presses Universitaires de France 1962.

Noll, W. (1967). Space-time structures in classical mechanics. In M. Bunge, Ed., *Delaware Seminar in the Foundations of Physics*, pp. 28–34. New York: Springer-Verlag.

Ockham, W. (ca. 1320). *Summa totius logicae*. In P. Boehner, Ed., *Ockham: Philosophical Writings*. Edinburgh: Nelson 1957.

Ostwald, W. (1902). *Vorlesungen über Naturphilosophie*. Leipzig: Veit & Comp.

Padulo, L., and M. Arbib (1974). *System Theory*. Philadelphia: W. B. Saunders.

Peirce, C. S. (1892–93). *Scientific Metaphysics*. Vol. VI of *Collected Papers*, C. Hartshorne and P. Weiss, Eds. Cambridge, Mass.: Harvard University Press 1935.

Peña, L. de la (1969). New formulation of stochastic theory and quantum mechanics. *J. Math. Phys.* **10**: 1620–1630.

Peña, L. de la, and A. M. Cetto (1975). Stochastic theory for quantum mechanical systems. *Foundations of Physics* **5**: 355–373.

Penney, R. (1965). On the dimensionality of the real world. *J. Math. Phys.* **6**: 1607–1611.

Penrose, R. (1971). Angular momentum: an approach to combinatorial spacetime. In T. Bastin, Ed., *Quantum Theory and Beyond*, pp. 151–180. Cambridge: University Press.

Pielou, E. C. (1969). *An Introduction to Mathematical Ecology*. New York: Wiley-Interscience.

Plantinga, A. (1974). *The Nature of Necessity*. New York: New York University Press.

Popper, K. R. (1957). The propensity interpretation of the calculus of probability, and the quantum theory. In S. Körner, Ed., *Observation and Interpretation*, pp. 65–70. London: Butterworths Scientific Publications.

Popper, K. R. (1963). The demarcation between science and metaphysics. In P. A. Schilpp, Ed., *The Philosophy of Rudolf Carnap*, pp. 877–881. La Salle, Ill.: Open Court.

Popper, K. R. (1968). On the theory of the objective mind. Repr. in *Objective Knowledge*, pp. 153–190. Oxford: Clarendon Press 1972.

Popper, K. R. (1974). Autobiography. In Schilpp, Ed. (1974), Bk. I, pp. 3–181.

Putnam, H. (1969). On properties. In N. Rescher, Ed., *Essays in Honor of Carl G. Hempel*, pp. 235–254. Dordrecht: Reidel.

Quine, W. V. (1954). Reduction to a dyadic predicate. Repr. in *Selected Logic Papers*, pp. 224–226. New York: Random House 1966.

Quine, W. V. (1963). *Set Theory and its Logic*. Cambridge, Mass.: The Belknap Press of Harvard University Press.

Quinton, A. (1973). *The Nature of Things*. London & Boston: Routledge & Kegan Paul.

Raggio, A. R. (1969). Was heisst 'Bedingungen der Möglichkeit'? *Kant Studien* **60**: 153–165.

Rashevsky, N. (1960). *Mathematical Biophysics*, 3rd ed. New York: Dover.

Regge, T. (1967). Review of Chew's book (Chew 1966). *Physics Today* **20**: No. 9: 82–87.

Reichenbach, H. (1928). *The Philosophy of Space and Time*. New York: Dover 1957.

Reichenbach, H. (1956). *The Direction of Time*. Berkeley: University of California Press.

Rescher, N. (1975). *A Theory of Possibility*. Oxford: Basil Blackwell.

Rescher, N. and A. Urquhart (1971). *Temporal Logic*. Wien, New York: Springer-Verlag.

Robb, A. A. (1914). *A Theory of Time and Space*. Cambridge: University Press.

Robb, A. A. (1921). *The Absolute Relations of Time and Space*. Cambridge: University Press.

Robb, A. A. (1936). *Geometry of Time and Space*. Cambridge: University Press.

Rosenblueth, A. (1970). *Mind and Brain: a Philosophy of Science*. Cambridge, Mass.: MIT Press.

Russell, B. (1900). *A Critical Exposition of the Philosophy of Leibniz*. Cambridge: University Press.

Russell, B. (1914). *Our Knowledge of the External World*. London: Allen & Unwin 1952.

Russell, B. (1918). The philosophy of logical atomism. *The Monist* **28**: 495–527; **29**: 32–63, 190–22, 345–380. Repr. in *Logic and Knowledge*, R. C. Marsh, Ed., pp. 175–281. London: Allen & Unwin 1956.

Russell, B. (1940). *An Inquiry into Meaning and Truth*. London: Allen & Unwin.

Russell, B. (1948). *Human Knowledge: its Scope and Limits*. London: Allen & Unwin.

Rutherford, D. E. (1965). *Introduction to Lattice Theory*. Edinburgh & London: Oliver & Boyd.

Ryle, G. (1949). *The Concept of Mind*. London: Hutchinson.

Schilpp, P. A., Ed. (1974). *The Philosophy of Karl Popper*, 2 vols. La Salle, Ill.: Open Court.

Schlick, M. (1963). *Space and Time in Contemporary Physics*. New York: Dover.

Scholz, H. (1941). *Metaphysik als strenge Wissenschaft*. Darmstadt: Wissenschaftliche Buchgesellschaft 1965.

Schreider, Ju. A. (1975). *Equality, Resemblance, and Order.* Transl. M. Greendlinger. Moscow: MIR Publishers.

Schulz, C. M. (1967). *You're you, Charlie Brown.* New York: Holt, Rinehart & Winston.

Scott, D. (1970). Advice on modal logic. In K. Lambert, Ed., *Philosophical Problems in Logic: Some Recent Developments.* Dordrecht: Reidel.

Sellars, W. (1963). *Science, Perception and Reality.* London: Routledge & Kegan Paul.

Settle, T. (1974). Induction and probability unfused. In Schilpp, Ed. 1974, Bk. II, pp. 697–749.

Shea, W. (1972). *Galileo's Intellectual Revolution.* London: Macmillan.

Sinnot, E. W. (1963). *The Problem of Organic Form.* New Haven: Yale University Press.

Smart, J. J. C. (1963). *Philosophy and Scientific Realism.* New York: Humanities Press.

Smart, J. J. C., Ed. (1964). *Problems of Space and Time.* New York: Macmillan.

Sneath, P. H. and R. R. Sokal (1973). *Numerical Taxonomy.* San Francisco & London: Freeman.

Sobociński, B. (1954–55). Studies in Lesniewski's mereology. *Polskie Towarzystwo Naukowe na Obezyźnie Yearbook*, London.

Stapp, H. P. (1971). S-matrix representation of quantum theory. *Phys. Rev.* D **3**: 1303–1320.

Stapp, H. P. (1977). Theory of reality. *Foundations of Physics*, forthcoming.

Stiehler, G. (1967). *Der dialektischer Widerspruch*, 2nd ed. Berlin: Akademie-Verlag.

Strauss, M. (1970). Intertheory relations. In P. Weingartner and G. Zecha, Ed., *Induction, Physics, and Ethics.* Dordrecht: Reidel.

Strawson, P. F. (1959). *Individuals: an Essay in Descriptive Metaphysics.* London: Methuen. Repr. Cambridge University Press 1964.

Suppes, P. (1970). *A Probabilistic Theory of Causality.* Acta Philosophica Fennica XXIV. Amsterdam: North-Holland.

Suppes, P. (1974). *Probabilistic Metaphysics*, 2 vols. Uppsala: Filosofiska Studier.

Tarski, A. (1927). Foundations of the geometry of solids. In *Logic, Semantics, Metamathematics*, pp. 24–29. Transl. J. H. Woodger. Oxford: Clarendon Press.

Tarski, A. (1944). The semantic conception of truth. *Phil. and Phenomenological Research* **4**: 341–375.

Tarski, A. (1954). Contributions to the theory of models, I, II. *Indagationes Mathematicae* **16**: 572–588.

Teilhard de Chardin, P. (1964). *The Phenomenon of Man.* New York: Harper.

Teller, P. (1975). Essential properties: some problems and conjectures. *J. Phil.* **72**: 233–248.

Terletskii, Ya. P. (1971). *Statistical Physics.* Amsterdam: North-Holland; New York: Elsevier.

Trautman, A. (1965). Foundations and current problems of general relativity. In A. Trautman, F. A. E. Pirani and H. Bondi, *Lectures on General Relativity I*, pp. 1–248. Englewood Cliffs: Prentice-Hall.

Tripathi, R. K. (1969). The central problem of Indian metaphysics. *Phil. East and West* **19**: 39–43.

Truesdell, C. (1974). A simple example of an initial value problem with any desired number of solutions. *Rendiconti dell'Istituto Lombardo*, Classe di Scienze (A) **108**: 301–304.

Ujomov, A. I. (1965). *Dinge, Eigenschaften und Relationen.* Berlin: Akademie-Verlag.

Vuillemin, J. (1971). *La logique et le monde sensible.* Paris: Flammarion.

Wallace, A. D. (1941). Separation spaces. *Ann. Math.* (2) **42**: 687–697.

Walsh, W. H. (1963). *Metaphysics.* London: Hutchinson.

Weiss, P. (1963). The cell as unit. *J. Theoret. Biology* **5**: 389–397.

Weyl, H. (1940). The evasive ghost of modality. In M. Farber, Ed., *Philosophical Essays in Memory of Edmund Husserl*, pp. 278–303. Cambridge, Mass.: Harvard University Press.

Weyl, H. (1949). *Philosophy of Mathematics and Natural Science.* Princeton: University Press.

Wheeler, J. A. (1962). *Geometrodynamics.* New York: Academic Press.

Whitehead, A. N. (1919). *An Enquiry Concerning the Principles of Natural Knowledge.* Cambridge: University Press.

Whitehead, A. N. (1920). *The Concept of Nature.* Repr. University of Michigan Press 1957.

Whitehead, A. N. (1929). *Process and Reality.* New York: Macmillan. Repr. 1969.

Whitrow, G. J. (1961). *The Natural Philosophy of Time.* London: Thomas Nelson. Paperback: Harper Torchbooks 1963.

Williams, D. C. (1937–38). The realistic interpretation of scientific sentences. *Erkenntnis* **7**: 169–178, 375–382.

Williams, D. C. (1953). On the elements of being. *Rev. Metaphys.* **7**: 3–18, 171–192.

Wilson, N. L. (1955). Property and description. *Phil. Rev.* **64**: 389–404.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus.* London: Routledge & Kegan Paul. Rev. ed. 1951.

Wittgenstein, L. (1969). *Über Gewissheit.* Ed. by G. E. M. Anscombe and G. H. von Wright. Oxford: Basil Blackwell.

Woodger, J. H. (1929). *Biological Principles.* London: Routledge & Kegan Paul.

Woodger, J. H. (1939). *The Technique of Theory Construction.* Chicago: University of Chicago Press.

Woodger, J. H. (1951). Science without properties. *Brit. J. Phil. Sci.* **2**: 193–216.

Yessenin-Volpin, A. S. (1970). The ultra-intuitionistic criticism and the antitraditional program for foundations of mathematics. In A. Kino, J. Myhill and R. E. Wesley, Eds., *Intuitionism and Proof Theory*, pp. 3–45. Amsterdam: North-Holland.

Yukawa, H. (1973). *Creativity and Intuition.* Transl. J. Bester. Tokyo: Kodansha International.

Zadeh, L. A. and C. A. Desoer (1963). *Linear System Theory; the State Space Approach.* New York: McGraw-Hill.

Zeeman, E. C. (1962). The topology of the brain and visual perception. In M. K. Fort Jr., Ed., *Topology of 3-Manifolds.* Englewood Cliffs, N.J.: Prentice-Hall.

INDEX OF NAMES

Agassi, Joseph 16, 334
Alembert, Jean le Rond d' 18
Alexander, H. G., Ed. 74, 334
Alexander, Samuel 279, 334
Allen, J. A. 153, 334
Anaximander 279
Aquinas, Thomas xiii, 4, 6, 269, 334
Arbib, Michael 121, 123, 340
Aris, Rutherford xv
Aristotle xiii, 4, 6, 26, 73, 103, 164, 167, 172, 174, 179, 183, 213, 254, 273, 275, 280, 284, 296, 310, 312, 327, 329, 334
Ashby, William Ross 150, 334
Augustine 280

Basri, S. A. 282, 295, 334
Bergmann, Gustav 168, 334
Bergmann, Hugo 299, 334
Bergson, 3, 334
Berkeley, George 139, 281
Bernal, John D. 18
Bernays, Paul 108, 334
Blalock, Hubert M., Jr. 95, 334
Blohm, Robert xv
Blokhinzev, D. I. 195
Bohm, David 274, 334
Bohr, Niels 272
Bolzano, Bernard 8, 60, 91, 99, 121, 173, 269, 334
Booij, H. L. 242, 334
Borel, Émile 207, 334
Born, Max 18
Bradley, Francis Herbert 70, 100, 334
Braudel, Fernand 252, 335
Brody, Thomas A. xv
Brown, Charlie 51
Bruck, R. H. 29, 335
Bunge, Eric 2
Bunge, Mario 2, 7, 15, 18, 49, 55, 60, 74, 131, 160, 191, 195, 254, 285, 299, 320, 322, 323, 335, 336

Carnap, Rudolf 50, 152, 183, 268, 282, 336
Castañeda, Héctor-Neri 103, 336
Cetto, Ana María 196, 340
Chaitin, G. J. 210, 336
Chew, Geoffrey 56, 336
Chisholm, Roderick, Ed. 3, 336
Chrysippus 172, 178, 213
Cicero 172
Clifford, William Kingdon 279
Cocchiarella, Nino B. 57, 203, 336
Collingwood, Robin George 3, 4, 16, 336
Condillac, Etienne 282
Costa de Beauregard, Olivier 269, 336
Cournot, Antoine-Augustin 213, 336
Cox, D. R. 208, 336
Cresswell, M. J. 167, 202, 336

Dantzig, D. van 280, 336
Darwin, Charles 18
Dehnen, Hans 266, 337
Democritus 269
DeWitt, Bryce S. 211, 337
DeWitt, Norman Wentworth 35, 52, 337
Descartes, René xiii, 6, 18, 26, 273, 280, 284, 294, 310, 327, 330
Desoer, Charles A. 123, 135, 343
Diodorus Chronos 164
Dobzhansky, Theodosius 18
Ducasse, Curt J. 69, 337

Einstein, Albert 18, 195, 305, 310, 315, 330, 337

Engels, Friedrich xiii, 5, 214, 255, 280, 337
Epicurus 46, 52, 213
Euler, Leonhard 18
Everett III, Hugh 211, 337

Falk, G. 104, 337
Faraday, Michael 18, 311
Feibleman, James 102, 337
Feigl, Herbert 184, 337
Feller, William 249, 337
Finkelstein, David 274, 337
Fitch, Frederic B. 112, 337
Fokker, A. D. 273, 337
Fréchet, Maurice 192, 337
Friend, J. W. 102, 337

Gaal, S. A. 291, 337
Galilei, Galileo 18, 310
García Máynez, Adalberto xv, 287, 336
García Sucre, Máximo xv
Garza, Tomás xv
Gell-Man, Murray 207, 337
Giere, Ronald 191, 337
Gonseth, Ferdinand 13, 337
Goodman, Nelson 39, 50, 53, 102, 121, 152, 268, 337, 338
Grünbaum, Adolf 269, 299, 337

Haldane, J. B. S. 18
Harary, Frank 70, 100, 337
Harré, Rom 16, 337
Hart, R. G. 211, 338
Hartnett, William E. xv
Harvey, D. 16, 338
Hasenjäger, Gisbert 13, 338
Hebb, Donald 11, 338
Hegel, Georg Wilhelm Friedrich 5, 6, 13, 51, 159, 255, 270, 338
Heidegger, Martin 3, 6, 14, 51, 159, 338
Heisenberg, Werner 18, 19, 338
Helmholtz, Hermann von 69, 338
Heraclitus 269
Hilbert, David 284, 338
Hintikka, Jaakko 202, 338
Hirsch, M. 133, 338
Hiż, Henry 104, 338
Hobbes, Thomas xiii, 213
Holbach, Paul Henri d' xiii
Hooker, Joseph Dalton 153, 338
Huber, Gerhard xv
Huber, Peter xv
Hughes, G. E. 167, 338
Hume, David 81, 173, 338
Huntington, Edward V. 289–290, 338
Husserl, Edmund 118

James, William 173
Jammer, Max 278, 338
Jensen, Uffe Juul 103, 338

Kálnay, Andrés J. xv
Kant, Immanuel 22, 213, 279, 329, 338
Klir, George, Ed. 5, 338
Kneale, Martha 172, 338
Kneale, William 172, 338
Körner, Stefan 16, 338
Koyré, Alexandre 14, 338
Kripke, Saul 202, 203, 338

Lakatos, Imre 16, 339
Lambert, Jean Henri 280, 339
Langford, C. H. 199, 339
Leibniz, Georg Wilhelm xiii, 6, 13, 18, 74, 90–91, 121, 202, 280, 339
Lejewski, Czeslaw 38, 338
Leonard, Henry S. 38, 50, 102
Łesniewski, Stanisław 5, 38, 50
Lévi-Strauss, Claude 275, 339
Lewes, George Henry 97, 339
Lewis, Clarence Irving 13, 102, 199, 203, 339
Lewis, David 203, 339
Ljapin, E. S. 27, 339
Lobachevsky, N. I. 280
Locke, John 67, 339
Lotze, Rudolf Hermann xiii, 329, 339
Lovejoy, Arthur A. 328, 339
Lucas, J. R. 282, 291
Luce, R. Duncan 243, 339
Lucretius 35, 52, 280, 339
Łukasiewicz, Jan 213, 339

Mach, Ernst 152, 280–282, 339
Marcus, Ruth Barcan 199, 339
Margenau, Henry 16, 263, 339
Maritain, Jacques 159
Marsh, R. C., Ed. 267, 341
Martin, Richard M. 50, 339
Maxwell, James Clerk 18, 204, 339
McGilvary, Evander Bradley 282, 340
Meinong, Alexius, 3, 5, 340
Mellor, D. H. 196, 340
Melvin, M. A. 207, 340
Mendelson, Elliott 107, 340
Mesarović, Mihajlo D. 5, 123, 340
Mises, Richard von 192
Montagu, William Pepperell 5, 340
Montague, Richard 174, 199, 340
Munitz, Milton, Ed. 15, 340

Naess, Arne 213, 340
Neumann, Johann von 272, 340
Newton, Isaac 18, 180, 183, 311, 332, 340
Nicod, Jean 282, 340
Noll, Walter 332, 340

Ockham, William of 104, 340
Oparin, A. I. 166
Ostwald, Wilhelm 241, 340

Padulo, Louis 121, 123, 340
Parmenides 2, 13
Peirce, Charles Sanders xiii, 4, 8, 51, 102, 160, 161, 174, 213, 340
Peña, Luis de la 196, 340
Penney, R. 295, 340
Penrose, Roger 280, 340
Piaget, Jean xiv
Pielou, E. C. 150, 340
Plancherel, M. 248
Planck, Max 194
Plato 26, 91, 100, 118, 157, 270
Poincaré, Henri 192
Popper, Karl R. 22, 34, 183, 192, 193, 210, 341
Priestley, Joseph 18

Quine, William van Orman 2, 54, 69, 157, 341
Quinton, Anthony 5, 341

Raggio, Andrés R. 180, 341
Regge, Tullio 56, 341
Reichenbach, Hans 299, 341
Rescher, Nicholas 172, 341
Riemann, Bernhard 280
Robb, A. A. 299, 308, 341
Rosenblueth, Arturo 16, 17, 341
Rosenthal, A. 248
Russell, Bertrand xiii, 9, 16, 53, 60, 90, 103, 267, 268, 273, 282, 302, 311, 329, 341
Rutherford, D. E. 40, 341
Ryle, Gilbert 184, 341

Salmerón, Fernando xv
Salt, David xv
Sangalli, Arturo A. L. xv, 72, 231, 336
Sartre, Jean-Paul 3, 159
Scholz, Heinrich 4, 13, 15, 341
Schreider, Juri A. 92, 342
Schrödinger, Erwin 18
Schulz, Charles M. 51, 342
Sellars, Wilfrid 164, 342
Settle, Tom 196, 342
Shea, William 94, 342
Sinnott, E. W. 103, 342
Smale, Stephen 133, 338
Smart, J. J. C. 164, 342
Smoluchowski, Marian von 192
Sneath, P. H. 86, 342
Sobociński, B. 38, 342
Sokal, R. R. 86, 342
Spinoza, Benedict xiii, 213, 214
Stapp, Henry 211, 274, 342
Stiehler, G. 101, 342
Strauss, Martin 194, 342
Strawson, Peter Frederick 2, 3, 23, 75, 152, 310, 342
Suppes, Patrick 15, 171, 210, 342

Takahara, Y. 123, 340
Tarski, Alfred 14–15, 38, 342

Teilhard de Chardin, Pierre 86, 342
Teller, Paul 96, 342
Terletskii, Ya. P. 205, 342
Torretti, Roberto xv
Trautman, A. 292, 342
Truesdell, Clifford 175, 342

Ujomov, A. I. 104, 343
Urquhart, Alasdair 172, 341

Vollmer, Gerhard xv
Vuillemin, Jules 282, 343

Wallace, A. D. 47, 285, 343
Weiss, Paul 103, 343
Weyl, Hermann 208, 269, 343
Wheeler, John Archibald 279, 343
Whitehead, Alfred North xiii, 4, 9, 58, 86, 102, 152, 268, 273, 281, 282, 291, 302, 343
Whitrow, George J. 297, 343
Williams, Donald C. 5, 343
Wilson, Neil L. 93, 343
Wittgenstein, Ludwig 54, 152, 268, 343
Wolff, Christian 202
Wolvekamp, H. P. 242, 334
Woodger, Joseph Henry 5, 16, 57, 102, 343

Yessenin-Volpin, A. S. 75, 343
Yukawa, Hideki 24, 343

Zadeh, Lofti 123, 135, 343
Zeeman, E. C. 92, 343

INDEX OF SUBJECTS

A-composition 47–48
Accident 94
Act 183
Action 258–260
Action at a distance 323
Action, nearby 323–326
Actualism 164, 211–213
Actuality 170–171
Aggregate 263–264
Aggregation 32, 114
Analysis, philosophical 9–10
Antecedence principle 320–323
Assemblage 43
Assembly theory 39–50
Association theory 26–36
Atom 36, 147
Atom aggregate 36–37
Atomism 35–36
  logical 268
Attribute 57–62

Basic thing 35–36
Bayesian interpretation 191–192
Black box theory 55–56
Bond 261
Bondage 261–262
Boolean algebra 49
Bulk 293

Canonical transformations 228–229
Capacity of a thing 135
Category, mathematical 224
  ontological 10–12
Causality 210–211, 320–323, 327
Causal relation 298, 326–327
Chain 243–244
Chance 194–199. *See also* Randomness
Change 215–275
  in spacetime 312–314
  net 221
  overall 219
  qualitative 220
  quantitative 220
  rate of 237–238
  serial 243–251
Changeability 218–219
Chaos 209–210
Chronotopics 279–283, 331–333
Circumstance 174–176
Class 140–143
Clock 301–302
Clustering 37–38
Collection, variable 236–237
Complexity 43, 56
Composite 29, 42
Composition 31, 43. *See also* *A*-composition
Concept 52–53
Connectibility 300–302, 315
Conservation 28, 34–35, 46
Construct 34, 52–53, 116–119, 157, 162–163, 219, 310
Contiguity, spatiotemporal 325–326
Continuity 239–240
Contradiction 73
Counterfactual 184–185
Coordinatization of states 234–235
Creation *ex nihilo* 254

Decomposition 42, 44, 46
Detachment relation 31, 42, 46–47
Dialectics 89
Dichotomization of predicates 69–71
Dimensionality of physical space 291–292, 295
Disposition 179–185
Dissimilarity 87–88

Duration 296–305
Dynamicism 268–270
Dynamics, Lagrangian 21

Effect 261
Emergence 97–98
Empiricism 282
Energetism 241
Energy 240–243
EPR paradox 259, 264
Equifinal 246
Equilibrium 249
Equivalence, nomological 147
Erlebnis 152
Essence 96
Essentialism 96
Event 217, 226
  composition 224, 231, 236, 284
  extent of 238–239
  lawful 229–230
  order 226, 235
  precedence 225–226
  space 222–226, 230–231
Existence 54, 155–162
  conceptual 157
  criteria 160–162
  in space and time 327–328
  of space and time 328–330
  predicate 155–156
  real 157
  virtual 159–160
Extensionalism 57, 100

Fact, 168–171, 267–268
Field 315
Filter, mathematical 141
  of classes of things 142–150
Finitism 43
Form 57–109. *See also* Property
Framework, ontological 12
Frequency interpretation 191
Functional schema 119–122

Genetics 131–132
Genus, natural 151
Geometry 277–278
  mathematical 277
  philosophical 278, 285–287
  physical 277–278
Gestalt 97

History of a thing 217, 255–260, 314, 319
Holism 97

Ideal, mathematical 141
  of classes of things 142–150
Idealism 112, 121
Identity law for things 74–75
Indentity of attributes 93–94
Ideology 24
Indiscernibility 90–92
Individual, substantial 26–56
Individuals, calculi of 50
Individuation 54, 58, 74, 111, 257
Infinitism 43
Infinity 121
Information process 271–272
Integration, degree of 190
Interaction 259
Interconnectedness 270–271
Interposition relation 284–285
Invariance 66, 137, 270, 316, 319–320
  nominal 221
Irreversibility 247–248, 318

Juxtaposition 39–50, 114–115

Kind 143–150
  natural 143, 146

Lattice 40–41
Law 77–80, 124, 173–176
  basic 151
  derived 151
  function 129, 136
  statement 128–131
Lawfulness 173–174, 252
Levels 47–48

Limiting factor 242
Logic 13–15
  modal 167–168, 199–202
Lorentz transformations 137

Materialism 152
Mathematics 13–15, 186
Measurement theory 272–273
Memory 247
Mereology 38
Metaphysics: *see* Ontology
Model 135
  kinematical 313
  mechanical 313
  set 202
  thing 119–123. *See also* Functional schema
Molecule 147
Monism 12
Monoid 27
  free 37
Moore graph 223
Motion 314

Naturalism 12
Necessity, conceptual 167–168
  factual 174–176
Neuron 131
Nominalism 57–58, 96, 121, 154–155
Nothingness 158–159
Novelty 98
Null individual 27–29, 35, 50–52

Object 117–118
Ontology xiii, 1–25, 118, 275, 278
  exact 7–8
  of science 15–22
  of technology 19–22
  regional 11
  scientific 7–12
  universal 11
Operationism 44
Opposite 88–89, 101

Parallelism, psychophysical 272
Part-whole relation 29–30, 43, 114
Pauli principle 263
Phase space 228–229
Phenomenalism 162
Physics 292
Place 293, 310–311
Platonism 2, 53, 58, 100, 102–108, 154–155
Pluralism 13
Polarity 101
Position in spacetime 309–311
Possibilism 211–214
Possibility 164–214
  conceptual 165–168
  criteria 177–179
  epistemic 165–166
  field 206
  logical 165–166
  real 134, 171–174
  theoretical 177–179
Possible worlds metaphysics 34, 202–204
Potentiality 198
  causal 182
  chance 197–198
Principium individuationis: *see* Individuation
Probability 185–194, 204–208, 210–211
  state space 187–190
Process 225–255
  continuous 244–245
  hereditary 247
  metaphysics 275
Propensity, causal 179–185
  chance 179, 190–199
Property 57–109
  atomic 84–85
  basic 84–85
  compatibility 73, 76
  concomitance 81–82
  conjunction 82
  formal 106
  general 62–71
  individual 63, 71–92
  intrinsic 65
  manifest 198–199

Property—*continued*
  mutual 65–69
  negative 60–61
  phenomenal 66–68
  precedence 80–82
  qualitative 68
  quantitative 68
  secondary 66–68
  stochastic 126
  substantial 71
  universal 83
  weight 94–96
Pseudoquestion 23

Quantifier 155
Quantum theory 123, 128, 132, 136, 138–139, 159–160, 194–196, 211–212, 263, 272–273, 326

Randomness 173–174, 208–210, 249
  degree of 209
Reaction 132, 245, 260
Realism, epistemological 272
  naive 61–62
Reality 161. *See also* World
  of properties 99–104
  of relations 287
  of space and time 328–330
Reduction 98. *See also* Vol. 4, Ch. 9
Reference frame 123–124, 232–235, 264–267
  state 232–237
Reification 58
Relativity theories 177, 299–300, 308, 332
Representation 59–60, 132, 136–138
  equivalent 136–138
  functional 228
Representative point 216
Resultant 97–98
Reversibility 247–248, 318
  degree of 248
Ring 49

Science xiii, 15–22, 55–56, 152, 190
Scope of a property 75–76, 140
Semantics 14–15. *See also* Vols. 1 and 2
Semantic hypotheses 280
Semiabstract theory 186
Semigroup 27
Semilattice 33
Separation between events 306–307
  between things 285–287
Set theory 107–108
Shape 293–294
Simple individual 29, 42
Simultaneity 85–90
Space 276–296
  of facts 171
  philosopher's 285–287
  physical 288–292
  views on 278–281, 330–333
Spacetime 305–314
Species 153–155. *See also* Kind
Spontaneity 257
Stability 249–250
Standard thing: *see* Reference frame
State 123–140, 169, 266, 283–284
  function 125–129, 232–233
  heterogeneity 188–189
  homogeneity 188–189
  precedence 298
  preparation 138–139
  space 131–139, 215–220, 226–237, 266
    lawful 133–134, 227–228
    transformation 227–228
  variable: *see* State function
Structuralism 275
Structure 121–122. *See also* Vol. 4, Ch. 7
Subjectivism 139, 191–192, 282
Substance 26–56
Superposition 39–50
Synthesis, philosophical 9–10
System 263–264. *See also* Vol. 4, Ch. 7
Systems theory 123–124

Theory, ontological 10–12
  scientific 177–178
    ontological background of 20
Thing 110–163

Time 296–305, 327–330
  arrow of 319
  local 303
  reversal 317–320
  travel 322
  universal 303
Tolerance space 92
Truth condition for probability statements 193

Unarization of predicates 69–71
Universal 104–108
Universality, degree of 105

Variety 150–151

World 30, 33, 34, 114, 152
  line 274
Worldly 30, 34

The companion of this book is Volume 4 of the

Treatise on Basic Philosophy
ONTOLOGY II

A World of Systems

Table of Contents

PREFACE

INTRODUCTION

7 System
8 Novelty
9 Biosystem
10 Psychosystem
11 Sociosystem
12 Synopsis
BIBLIOGRAPHY
INDEX OF NAMES
INDEX OF SUBJECTS

furnitureofworld00bung

furnitureofworld00bung

furnitureofworld00bung

The present volume offers a new ontology. This system is materialist (though not reductionist) and dynamicist (though neither dialectical nor Whiteheadian). It is in tune with contemporary science and built with the help of some formal tools. It elucidates and systematizes, among others, the concepts of substance and property, thing and system, possibility and actuality, probability and law, state and event, and space and time. Many of the axioms and theorems of the system are found to be presupposed or else implied by contemporary scientific research. Therefore the work constitutes a bridge between the metaphysical tradition and contemporary philosophy of science. There are many examples and bibliographical references.

Dr Mario Bunge has been a professor of theoretical physics, and later on of philosophy, in his native Argentina and in the U.S.A. He has held visiting positions at the Universities of Pennsylvania, Texas, Temple, Delaware, Freiburg, and Aarhus, as well as at the E.T.H. in Zürich. He is currently professor of philosophy and head of the Foundations and Philosophy of Science Unit at McGill University. He has been awarded Guggenheim, Alexander von Humboldt and Ernesto Santamarina fellowships, as well as Killam grants. He has written a dozen books and edited seven volumes, one journal of philosophy and another of physics, and has founded three philosophical societies. His best known works are *Causality* (available in eight languages), *Scientific Research* (in two volumes), and *Foundations of Physics.* His latest publications are *Philosophy of Physics* and *Method, Model and Matter*. His most recent editorial jobs, *Problems in the Foundations of Physics*, *Exact Philosophy: Problems, Tools, Goals*, and *The Methodological Unity of Science.*

Audience: Philosophy teachers and graduate students, natural and social scientists, and systems scientists. This book can be used as a textbook for courses on metaphysics or on exact philosophy, and as supplementary reading for courses on philosophy of science, foundations of science, or systems theory.

D. REIDEL PUBLISHING COMPANY
DORDRECHT-HOLLAND / BOSTON-U.S.A.

TREATISE ON BASIC PHILOSOPHY

Volume 4

ONTOLOGY II: A WORLD OF SYSTEMS

TREATISE ON BASIC PHILOSOPHY

1

SEMANTICS I *Sense and Reference*

2

SEMANTICS II *Interpretation and Truth*

3

ONTOLOGY I *The Furniture of the World*

4

ONTOLOGY II *A World of Systems*

5

EPISTEMOLOGY I *The Strategy of Knowing*

6

EPISTEMOLOGY II *Philosophy of Science*

7

ETHICS *The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 4

Ontology II:

A WORLD OF SYSTEMS

D. REIDEL PUBLISHING COMPANY
DORDRECHT : HOLLAND / BOSTON : U.S.A.
LONDON : ENGLAND

Library of Congress Cataloging in Publication Data (Revised)

Bunge, Mario Augusto
Ontology.

(His Treatise on basic philosophy ; v. 3–4)
Includes bibliographies and indexes.
CONTENTS: 1. The furniture of the world. 2. A world of systems.
1. Ontology–Collected works. I. Title.
BD311.B84 111 77-1750
ISBN-13: 978-90-277-0945-5 e-ISBN-13: 978-94-009-9392-1
DOI: 10.1007/978-94-009-9392-1

Published by D. Reidel Publishing Company,
P.O. Box 17, Dordrecht, Holland

Sold and distributed in the U.S.A., Canada, and Mexico
by D. Reidel Publishing Company, Inc.
Lincoln Building, 160 Old Derby Street, Hingham,
Mass. 02043, U.S.A.

All Rights Reserved
Copyright © 1979 by D. Reidel Publishing Company, Dordrecht, Holland
Softcover reprint of the hardcover 1st Edition 1979
No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any informational storage and
retrieval system, without written permission from the copyright owner

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and of right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient – except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can

become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: "Do your own thing. Your reward will be doing it, your punishment having done it".

CONTENTS OF *ONTOLOGY II*

GENERAL PREFACE TO THE *TREATISE* v

PREFACE TO *ONTOLOGY II* xiii

ACKNOWLEDGMENTS xv

SPECIAL SYMBOLS xvi

1. SYSTEM 1
   1. Basic Concepts 3
      1.1. Aggregate and System 3
      1.2. Concrete System: Definition 5
      1.3. More of the Same 8
      1.4. Subsystem 11
      1.5. Level 13
      1.6. Systems Association 14
      1.7. Other Kinds of System: Property and Functional 15
      1.8. Concluding Remarks 15

   2. System Representations 16
      2.1. Coupling Graphs and Matrices 16
      2.2. The State Space Representation 20

   3. Basic Assumptions 24
      3.1. Structural Matters 24
      3.2. Assembly and Emergence 27
      3.3. Selection 30
      3.4. Evolution 33

   4. Systemicity 35
      4.1. Integration, Cohesion, Coordination 35
      4.2. Holism, Atomism, Systemism 39

   5. Concluding Remarks 43

2. CHEMISM 45
   1. Chemical System 46
      1.1. Atom 46

1.2. Molecule 48
1.3. Chemical System 52
1.4. Chemical Reaction 54
1.5. Chemical Control 59

2. Biochemical System 61
2.1. Biomolecule 61
2.2. DNA Replication, and RNA and Protein Synthesis 67
2.3. Biochemical System 71
2.4. Concluding Remarks 73

3. LIFE 75
1. From Chemism to Life 76
1.1. Self-Organization of Biochemical Systems 76
1.2. Definition of a Biosystem 79
1.3. Cell and Organism, Biospecies and Biopopulation 82
1.4. Biolevel 84
1.5. Concluding Remarks 85

2. Biofunction 86
2.1. Health and Death 86
2.2. Function and Value 88
2.3. Biocontrol 92
2.4. Development 95
2.5. Reproduction and Heredity 98
2.6. Development, Heredity, and the Metaphysical Tradition 101

3. Evolution 103
3.1. Adaptation 103
3.2. Adaptation Mechanism 105
3.3. Evolution and Coevolution: Basic Ideas 107
3.4. A Schema of the Evolutionary Process 108
3.5. The Underlying Mechanism 111
3.6. Bioprogress 115

4. Concluding Remarks 118

4. MIND 124
1. Central Nervous System 125
1.1. Philosophical Background 125
1.2. Neural Units 127
1.3. Initial Definitions 129
1.4. Basic Assumptions 132

2. Brain States 134

2.1. Brain Functions 134
2.2. Mental States and Processes 138
2.3. Psychosomatic Interaction 141
2.4. Location of the Mind 144
2.5. Mentalist Predicates 146

3. Sensation to Valuation 150
3.1. Sensing and Perceiving 150
3.2. Mapping Body and Environment 153
3.3. Behavior: Definitions and Principles 156

4. Recall to Knowledge 161
4.1. Memory and Learning 161
4.2. Expectation and Purpose 163
4.3. Thinking 164
4.4. Cognition and Decision 166
4.5. Creativity 167

5. Self to Society 170
5.1. Awareness, Consciousness, Will 170
5.2. Person and Self 173
5.3. Social Behavior 175
5.4. The Root of Social Cohesion 177
5.5. Communication 180
5.6. Protoeconomy, Protoculture, Protopolity 182

6. Concluding Remarks 183

5. SOCIETY 186
1. Human Society 187
1.1. Definitions 187
1.2. Subsystem, Supersystem. Specific Function 191
1.3. Institution 193

2. Social Subsystems and Supersystems 196
2.1. Work 196
2.2. Labor 198
2.3. Cultural and Managerial Work 200
2.4. The Main Subsystems of a Human Society 203
2.5. The Supersystems: Nation and World 205

3. Economy, Culture, and Polity 208
3.1. The Economic System 208
3.2. The Cultural System 211
3.3. The Political System 215

4. Social Structure 221
   4.1. Root Social Relations 221
   4.2. Social Structure 226
   4.3. Differentiation and Cohesion 229
   4.4. Social Relations and Structures: Real? 234

5. Social Change 235
   5.1. The Social Structure Approach 235
   5.2. History 237

6. Concluding Remarks 241

6. A SYSTEMIC WORLD VIEW 245
   6.1. A World of Systems 245
   6.2. System Genera 246
   6.3. Novelty Sources 247
   6.4. Emergence 249
   6.5. Systemism Supersedes Atomism and Holism 250
   6.6. Synopsis 251

APPENDIX A. SYSTEM MODELS 253
1. Input–Output Models 253
   1.1. The Black Box 253
   1.2. Connecting Black Boxes 255
   1.3. Control System 257
   1.4. Stability and Breakdown 260

2. Grey Box Models 262
   2.1. Generalities 262
   2.2. Deterministic Automata 264
   2.3. Probabilistic Automata 268
   2.4. Information Systems 270

APPENDIX B. CHANGE MODELS 273
1. Kinematical Models 273
   1.1. Global Kinematics 273
   1.2. Analytical Kinematics 274
   1.3. Balance Equations 275
   1.4. Lagrangian Framework 277
   1.5. Kinematical Analogy 280

2. Dynamical Models 281
   2.1. Generalities 281
   2.2. Formalities 282

2.3. The Pervasiveness of Cooperation and Competition 283
2.4. The Dynamics of Competitive–Cooperative Processes 285

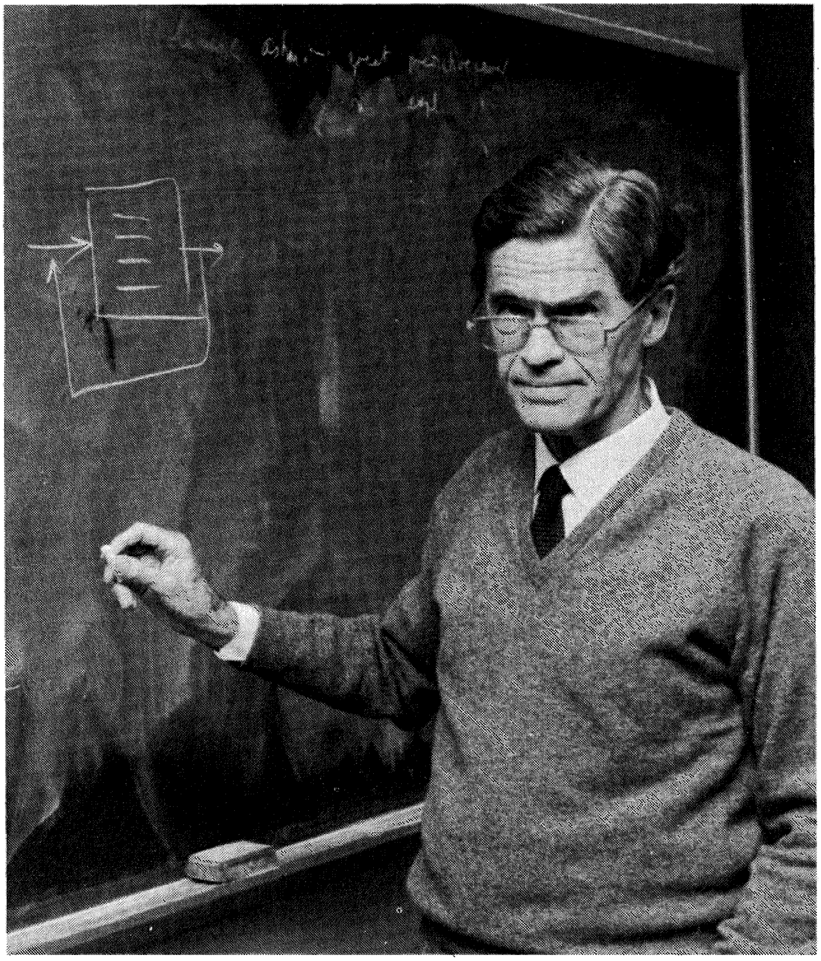
3. Qualitative Change Models 287
3.1. Kinematical: Birth and Death Operators 287
3.2. Dynamical: Random Hits 289

BIBLIOGRAPHY 292

INDEX OF NAMES 301

INDEX OF SUBJECTS 305

CONTENTS OF *ONTOLOGY I* (V. 3) 309



MARIO BUNGE

PREFACE TO *ONTOLOGY II*

This volume continues and concludes the task begun in Part I, titled *The Furniture of the World* – namely the building of an exact and systematic ontology consistent with contemporary science. However, it can be read independently by anyone willing to take for granted the basic notions analyzed and systematized in the companion volume, namely those of substance, property, thing, possibility, change, space, and time.

The three main themes of this book are wholeness (or systemicity), variety, and change. These three notions are analyzed and systematized, and they occur in some of the main assumptions of our ontology. One of these hypotheses is that the universe is not a heap of things but a thing composed of interconnected things – i.e. a system. This supersystem is composed of subsystems of various kinds: physical, biological, social, etc. Only physical systems may be composed of things that are not themselves systems, such as elementary particles and field quanta. However, even nonsystems are components of some system or other, and every system but the universe is a subsystem of some system: there are no strays. Ours is, in sum, a world of interconnected systems. Moreover it is the only one.

Another postulate of this system of ontology is that concrete systems are not all alike except insofar as they are systems and therefore tractable with the help of a unifying systems-theoretic framework. There are different kinds of system and each is characterized by its own peculiar properties and laws. Surely we sometimes succeed in accounting for the emergence and the history of a system in terms of its composition, environment, and structure. Nevertheless, explaining need not be explaining away: explained systems are not heaps, explained emergence is no mere resultant, and explained novelty is not old. Systemicity, emergence, and qualitative novelty and variety are as genuine as they are capable of being explained. Far from being incompatible with reason, wholeness and emergence can be understood.

A third major thesis of this work is that no system, except for the world as a whole, lasts forever. Systems get assembled, change, and break down. If natural, systems emerge as a result of self-assembly processes – often from debris of former systems. Even modest accretion processes can ensue in systems possessing emergent properties. Order can thus emerge from randomness, systems from physical precursors, living systems from nonliving ones, and so on. (Entropy need not increase in open systems.)

All three theses are by now common knowledge or nearly so. Now they – jointly with many others – have become part and parcel of a science-oriented ontological system couched in a fairly exact language. Thus the novelty of this system resides sometimes in its components, and at other times in their organization.

This volume covers the following ground. Chapter 1 defines the notions of system and subsystem, and emergence and level, and it lays down a handful of laws supposed to be shared by systems of all kinds. Chapter 2 defines the notions of chemical system and biochemical system, regarding them as emergent relative to the physical level. Chapter 3 attempts to catch the elusive notion of life, and tackles the problems of biovalue, health, adaptation, and bioprogress. Chapter 4 is devoted to the mind-body problem from a point of view that combines the systemic approach with the psychobiological one. (Individual neurons cannot think; but certain systems of neurons, if connected in special plastic ways, can.) Chapter 5 defines the concept of a sociosystem and analyzes every society into four main subsystems – the kinship system, the economy, the culture, and the polity. Chapter 6 presents the world view that emerges from the preceding chapters – one which is naturalistic (but not physicalist), systemic (but not holistic), pluralistic (with respect to properties, not substances), and dynamicist. Finally, the two appendices review some mathematical models of systems and of change which, because of their generality, belong to ontology as well as to science. These appendices are helpful but not essential for the understanding of the text.

ACKNOWLEDGMENTS

I am grateful for the advice and criticism offered by Professors Lina Bettucci (Departamento de Biología, Universidad Autónoma Metropolitana, México), Ernesto M. Bravo (Instituto Superior de Medicina, La Habana), A. Brito da Cunha (Instituto de Biociências, Universidade de São Paulo), Sir Francis Crick (The Salk Institute), Bernardo Dubrovsky (Department of Psychiatry, McGill University), Máximo García-Sucre (Sección Física, Instituto Venezolano de Investigaciones Científicas), Bernd Küppers (Max-Planck-Institut für Biophysikalische Chemie, Göttingen), Rodolfo Llinás (Department of Physiology and Biophysics, New York University Medical Center), Roger Palfree (Department of Biology, McGill University), Rafael Pérez Pascual (Instituto de Física, Universidad Nacional Autónoma de México), Osvaldo A. Reig (Departamento de Estudios Ambientales, Universidad Simón Bolívar, Caracas), Daniel Seni (Institut d'Urbanisme, Université de Montréal), and René Zayan (Centre de psychologie expérimentale et comparée, Université de Louvain). I am also indebted to my former research assistant Robert Blohm for suggesting a number of corrections. My former students David Conter, Adolfo García de la Sienra, Gordon Graham, and John Hawthorne made several remarks for which I am grateful. I am thankful to Dr Ricardo Peralta Fabi for drawing all the diagrams in the book. Last, but not least, I thank Professors D. R. Axelrad and Walter Hitschfeld, as well as the Canada Council and the McGill Faculty of Graduate Studies and Research, for supporting my research.

MARIO BUNGE

SPECIAL SYMBOLS

$\varnothing$	the *empty set*
$a \in A$	individual a *belongs* to set A
$A \subset B$	set A is *included* in set B
$A \cup B$	the set of objects in A *or* in B
$A \cap B$	the set of objects in A *and* in B
$A - B$	the set of objects *in* A *but not in* B
$A \triangle B$	the set of objects in A *or* in B but *not in both*
$\langle a, b\rangle$	the *ordered pair* of a and b
$A \times B$	the *cartesian product* of A and B
$\lvert A\rvert$	the *cardinality* (numerosity) of A
$a \dotplus b$	the *association* of things a and $b = \mathit{sup}\ \{a, b\}$
$\mathscr{C}(x)$	the *composition* of system x
$\mathscr{E}(x)$	the *environment* of system x
$\mathscr{S}(x)$	the *structure* of system x
$f : A \to B$	*function* f maps set A into set B
$f(x)$	the *value* of function f at x
$\mathbb{F} = \langle F_1, F_2, \ldots, F_n, \ldots\rangle$	*state function* for some system
$h(x, \tau)$	the *history* of thing x during time interval τ
iff	*if and only if* (necessary and sufficient condition)
$\mathbb{L}(x)$	the *set of laws* of thing x
$\mathbb{N}$	the set of *natural numbers*
$x \sqsubset y$	thing x is a *part* of thing y
$x \prec y$	system x is a *subsystem* of system y
2^S	the *power set* (family of subsets) of set S
Pr	*probability* function
$\mathbb{R}$	the *real line*
$\mathbb{R}^+$	the set of *non-negative real numbers*
$S_{\mathrm{L}}(x)$	*lawful state space* of thing x
T	set of *time* instants

CHAPTER 1

SYSTEM

Every science studies systems of some kind, whether natural (physical, chemical, biological, or social) or artificial (technical). Moreover most sciences study nothing but systems. Thus biology studies biosystems, sociology sociosystems, and technology technosystems. Physics seems to be the only science that investigates not only systems, such as atoms and large scale fields, but also putatively simple or elementary things such as electrons and photons. Even so, physicists acknowledge that every such basic thing is a component of some system or other

Until recently every species of system was studied separately. About four decades ago a number of specialists joined efforts to launch various cross-disciplinary ventures, such as operations research and cybernetics. Their success suggested to some workers that a unified approach to problems in various fields was possible. They pointed out that (*a*) there are some concepts and structural principles that seem to hold for systems of many kinds, and (*b*) there are some modeling strategies – in particular the state space approach – that seem to work everywhere.

The discipline that purports to develop such a unified approach is often called 'general systems theory' (Bertalanffy, 1950, 1958; Boulding, 1956). Paradoxically enough, this is not a single theory but a whole set of theories – automata theory, linear systems theory, control theory, network theory, general Lagrangian dynamics, etc. – unified by a philosophical framework (Bunge, 1974c, 1977c). We shall call *systemics* this set of theories that focus on the structural characteristics of systems and can therefore cross the largely artificial barriers between disciplines.

Systemics has two related motivations, one cognitive and one practical. The cognitive or theoretical rationale of systemics is, of course, the wish to discover similarities among systems of all kinds despite their specific differences – e.g. between body temperature control systems and furnace thermostats. The practical motivation for systemics is the need to cope with the huge and many-sided systems characteristic of industrial societies – such as communications networks, factories, hospitals, and armies. This complexity, in particular the variety of components of such systems, violates the traditional borders among disciplines and calls for a cross-disciplinary approach.

Note the differences between the standard scientist, engineer or social scientist on the one hand, and the systems "specialist" (actually a generalist) on the other. Whereas the former do or apply some particular science, the systemics expert de-emphasizes the physics (chemistry, biology, or sociology) of his systems, focusing instead on their structure and behavior. Moreover he is interested particularly in duplicating or imitating (modeling or simulating) the behavior of any given system (e.g. a person) by one of a different kind (e.g. a pattern recognition automaton). This holds not only for the mathematician who takes systemics as an honorable pretext for playing with abstract structures but without any serious concern for practical problems in engineering or management: it also holds for the systemist intent on solving practical problems, such as modeling and simulating a grazing land system or a university.

The method employed by the system theorist is mathematical modeling and the experimental (or at least the computer) testing of system models. Both are of course part of the scientific method. What is peculiar to the way the systemics expert proceeds is that, far from incorporating any specific (e.g. chemical) laws into his model, he aims at building a black box, a grey box, or a kinematical model free from details concerning the materials composing the system, and noncommittal enough to cover some of the global aspects of the organization and behavior of the system on some of its levels. The scientific method is thus taken for granted: what is emphasized is the general or cross disciplinary approach in contrast to the specific or disciplinary one. In other words, the systemics expert is a jack of all trades – a quasi-philosopher if not a full blown one.

Systemics is not quite the same as *systems analysis*, a much advertised, ill defined, and sometimes controverted thing. Systems analysis too, when serious, uses the scientific method but, unlike systemics, it is not particularly interested in de-emphasizing the peculiarities of the components of the system concerned. What it does emphasize is that, because it studies many-sided and multi-level systems – such as ecosystems and transportation systems – it must adopt various points of view on different levels. For example, hospitals are not just buildings with medical equipment but social systems as well – whose components include medical personnel and patients – and moreover subsystems of a larger social system, namely a health-care system, which is in turn a subsystem of a society. The novelty of systems analysis resides less in its methods than in the objects it studies, namely complex man-artifact systems never before approached in a scientific manner. Unlike systemics, systems analysis is hardly interested in building extremely general models: it aims instead at drawing flow charts, network diagrams, and occasionally specific mathematical models accounting if possible not just for the structure and kinematics of the system but also for its dynamics, and thus enabling one to understand how it operates and malfunctions, hence how it can be repaired. (For a hilarious account of systems antics, see Gall (1977).)

Systemics, or general system theory, is a field of scientific and technological research and one of considerable interest to philosophy. Because of its generality it has a sizable overlap with ontology or metaphysics construed in the traditional, pre-Hegelian sense as well as in our own sense of scientific ontology (Bunge, 1973a, 1977a). Both sytemics experts and ontologists are interested in the properties common to all systems irrespective of their particular constitution, and both are intrigued by the peculiarities of extremely general theories, which are methodologically quite different from specific theories (Bunge, 1973a, 1977c).

The main differences between systemics and ontology seem to be these: (*a*) while systems theorists take certain concepts for granted – e.g., those of property, possibility, change, and time – ontologists take nothing for granted except logic and mathematics; (*b*) while systems theorists are often interested in the details of the couplings of the components of a system, ontologists seldom are; (*c*) while systems theorists focus their attention on input–output models of systems that are largely at the mercy of their environment, ontologists are interested in free systems as well (in which respect they do not differ from physicists); (*d*) while systems theorists are mainly interested in deterministic (or rather nonstochastic) models – partly because theirs are large scale things – ontologists are also interested in stochastic ones; and (*e*) while some systems theorists focus their attention on the search for analogies among systems of different kinds, and particularly on different levels, ontologists are primarily interested in analyzing and systematizing concepts referring to all kinds of system.

In this chapter we shall propose a number of definitions and principles concerning concrete systems in general. These ideas will be used in succeeding chapters, where certain system genera will be studied. Details on mathematical models of systems are found in the two appendices.

1. BASIC CONCEPTS

1.1. *Aggregate and System*

An *aggregate* or assemblage is a collection of items not held together by bonds, and therefore lacks integrity or unity. Aggregates can be either conceptual or concrete (material). A conceptual aggregate is a set. (But not every set is a conceptual aggregate: a set equipped with a structure is a conceptual system.) A concrete or material aggregate, on the other hand, is a compound thing, the components of which are not coupled, linked, connected, or bonded, such as a field constituted by two superposed fields, a celestial constellation, and a random sample of a biological population.

Because the components of an aggregate do not interact – or do not interact appreciably – the behavior of each is independent of the behavior of the others. Consequently the history of the aggregate is the union of the histories of its members. On the other hand the components of a concrete system are linked, whence the history of the whole differs from the union of the histories of its parts. We shall take the last statement to be an accurate version of the fuzzy slogan of holistic metaphysics, namely *The whole is greater than the sum of its parts.* But we shall go far beyond this characterization of wholeness or systemicity. To this end we shall make use of a few elementary mathematical concepts as well as of a number of common notions – such as those of thing, property, and time – that have been clarified in our companion volume (Bunge, 1977a).

A system, then, is a complex object, the components of which are interrelated rather than loose. If the components are conceptual, so is the system; if they are concrete or material, then they constitute a concrete or material system. A theory is a conceptual system, a school a concrete system of the social kind. These are the only system kingdoms we recognize: conceptual and concrete. We have no use for mixed systems, such as Popper's "world 3", allegedly composed of conceptual objects, such as theories, as well as concrete objects, such as books (Popper, 1968; Popper and Eccles, 1977). We do not because, in order to be able to speak of the association or combination of two items, we must specify the association bond or operation. And, while mathematical theories specify the way conceptual items combine, and ontological and scientific theories take care of the combination of concrete items, no known theory specifies the manner whereby conceptual items could combine with concrete ones – and no experience suggests that such hybrids exist.

Whatever its kingdom – conceptual or concrete – a system may be said to have a definite composition, a definite environment, and a definite structure. The composition of a system is the set of its components; the environment, the set of items with which it is connected; and the structure, the relations among its components as well as among these and the environment. For example, a theory is composed of propositions or statements; its environment is the body of knowledge to which it belongs (e.g. algebra or ecology); and its structure is the entailment or logical consequence relation. The merger of these three items is a propositional system, i.e. a system $\mathcal{T}$ composed of a set P of propositions, embedded in a certain conceptual body B, and glued together by the relation $\vdash$ of entailment: in short, $\mathcal{T} = \langle P, B, \vdash \rangle$. And the composition of a school is the union of its staff and pupils; the environment is the natural and social milieu, and the structure consists of the relations of teaching and learning, managed and being managed, and others. The environment must be included in the description of a system because the behavior of the latter depends critically on the nature of its milieu. But of course in the case of the universe the environment is empty, and so it is in the case of the important fiction known as the free particle (or field).

One way of characterizing the general concept of a system is this. Let T be a nonempty set. Then the ordered triple $\sigma = \langle C, E, S \rangle$ is (or represents) a *system over* T iff C and E are mutually disjoint subsets of T (i.e. $C \cap E = \varnothing$), and S is a nonempty set of relations on the union of C and E. The system is conceptual if T is a set of conceptual items, and concrete (or material) if $T \subseteq \Theta$ is a set of concrete entities, i.e. things. However, the preceding is not a definition proper, because it does not tell us what exactly is the membership of the coordinates C, E and S of the ordered triple. We must therefore define the notions of composition, environment, and structure of a thing.

1.2. *Concrete System: Definition*

Let us start by defining the composition of a system. A social system is a set of socially linked animals. The brains of such individuals are parts of the latter but do not qualify as members or components of a social system because they do not enter independently into social relations: only entire animals can hold social relations. In other words, the composition of a social system is not the collection of its parts but just the set of its atoms, i.e. those parts that are socially connectible. This particular notion of composition is that of atomic composition or A-composition for short. It will be defined thus: The *A-composition* (or *composition at the A level*) of a thing x is the set of parts of x that belong to A. In symbols: let $A \subseteq \Theta$ be a class of things and let x be a thing (i.e. $x \in \Theta$). Then the (absolute) *composition* of x is the set of its parts, i.e.

$$\mathscr{C}(x) = \{y \in \Theta \mid y \sqsubset x\},$$

where '$y \sqsubset x$' designates "y is a part of x". And the *A-composition* of x is the set of its A-parts:

$$\mathscr{C}_A(x) = \mathscr{C}(x) \cap A = \{y \in A \mid y \sqsubset x\}.$$

Let us introduce next the concept of link, connection, or coupling among the components of a thing. We must distinguish between a mere relation, such as that of being older, and a connection, such as that of exerting pressure. Unlike a mere relation, a connection makes some difference to its relata. That is, two things are connected just in case at least one of them acts upon the other – where the action need not consist in eventuating something but may consist in either cutting out or opening up certain possibilities.

In turn, we say that one thing *acts* upon another if it modifies the latter's behavior line, or trajectory, or history. The acting of thing a on thing b is symbolized

$$a \rhd b.$$

If a thing acts upon another, and the latter does not react back, the former is called the *agent* and the latter the *patient*. If neither action nor reaction are nil, the things are said to *interact*. Finally, two things are *connected* (or *coupled*, or *linked*, or *bonded*) if at least one of them acts on the other. The *bondage* of a set $A \subseteq \Theta$ of things is the set $\mathbb{B}_A$ of bonds (or couplings or links or connections) among them. So, the total set of relations among the components of a complex entity may be decomposed into its bondage $\mathbb{B}_A$ and the set $\bar{\mathbb{B}}_A$ of nonbonding relations.

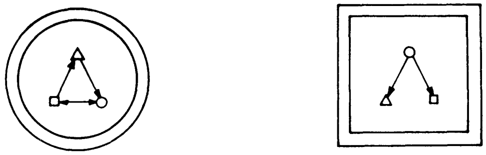
We can now introduce the notion of the *A-environment* of a thing x with A-composition $\mathscr{C}_A(x)$. It will be defined as the set of all things, other than those in $\mathscr{C}_A(x)$, that act on or are acted upon the latter:

$$\mathscr{E}_A(x) = \{y \in \Theta \mid \neg(y \in \mathscr{C}_A(x)) \;\&\; (\exists z)(z \sqsubset x \;\&\; (y \rhd z \vee z \rhd y))\}.$$

Finally the *structure* of a thing will be defined as the set of all the relations among the thing's components as well as among these and the things in the thing's environment.

We now have all we need to define the notion of a concrete system:

DEFINITION 1.1 An object is a *concrete system* iff it is composed of at least two different connected things.



(a) (b)

Fig. 1.1. Two systems with the same composition but different structures and environments.

Example A molecule, a coral reef, a family and a factory are systems. On the other hand a set of states of a thing and a collection of events, even if ordered, are not concrete systems. Symbol: Σ.

And now the three characteristics of any system:

DEFINITION 1.2 Let $\sigma \in \Sigma$ be a concrete system and $A \subset \Theta$ a class of things. Then

(i) the *A-composition* of σ at a given time t is the set of its A-parts at t:

$$\mathscr{C}_A(\sigma, t) = \{x \in A \mid x \sqsubset \sigma\};$$

(ii) the *A-environment* of σ at time t is the set of all things of kind A, not components of σ, that act or are acted on by components of σ at t:

$$\mathscr{E}_A(\sigma, t) = \{x \in A \mid x \notin \mathscr{C}_A(\sigma, t) \;\&\; (\exists y)(y \in \mathscr{C}_A(\sigma, t) \;\&\; (x \rhd y \vee y \rhd x));$$

(iii) the *A-structure* (or *organization*) of σ at time t is the set of relations, in particular bonds, among the components of σ, and among them and the things in the environment of σ, at t:

$$\mathscr{S}_A(\sigma, t) = \{R_i \in \mathbb{B}_A(\sigma, t) \cup \bar{\mathbb{B}}_A(\sigma, t) \mid \mathbb{B}_A(\sigma, t) \neq \emptyset \;\&\; 1 \leq i \leq n,\}$$

where $\mathbb{B}_A(\sigma, t)$ is the set of bonding relations, and $\bar{\mathbb{B}}_A(\sigma, t)$ that of non-bonding relations, defined on $\mathscr{C}_A(\sigma, t) \cup \mathscr{E}_A(\sigma, t)$.

Example The simplest possible system is one composed of two connected things, a and b, in an environment lumped into a single thing c. That is, $\mathscr{C}(\sigma) = \{a, b\}$, $\mathscr{E}(\sigma) = \{c\}$. This system can have either of the following internal structureses: $a \rhd b$, $b \rhd a$, or $a \bowtie b$: see Figure 1.2. (These are the conceivable internal structures. But some of them may not be nomologically possible, let alone technically feasible or even desirable.) As for

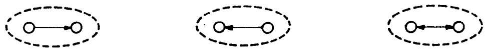


Fig. 1.2. A two component system with three possible internal structures.

the external structures, they can be any of these or their unions: $\{a \rhd c\}$, $\{b \rhd c\}$, $\{c \rhd a\}$, $\{c \rhd b\}$.

An exhaustive knowledge of a system would comprise the following items: (*a*) the composition, the environment and the structure of the system; (*b*) the history of the system (particularly if this is a biosystem or a sociosystem), and (*c*) the laws of the system. Such complete knowledge is seldom attainable, particularly with reference to complex systems. But in order to be able to speak of systems at all we should know at least their composition, their environment and their structure. Thus we can say that the model constituted by the ordered triple

$$s_A(\sigma, t) = \langle \mathscr{C}_A(\sigma, t), \mathscr{E}_A(\sigma, t), \mathscr{S}_A(\sigma, t)\rangle$$

is the *minimal model* of system σ at time t. Obviously, this qualitative model will not suffice for quantitative purposes, such as predicting the rate of formation or breakdown of a system. We shall therefore supplement the above minimal model with a quantitative model to be introduced in Sec. 2.2. However, before doing so we shall use the minimal model to clarify a number of issues that are often obscure in the literature on systems.

1.3. *More of the Same*

Before continuing our study of systems and their models we must make sure that the concept of a concrete system is not idle – i.e. that some things are systems while others are not. That some things are not systems follows from the assumption that there are basic or elementary things, i.e. things without parts (Vol. 3, Postulate 1.4). And that other things are systems follows from the ontological hypothesis that every thing – except the universe as a whole – acts on, and is acted upon by, other things (Vol. 3, Postulate 5.10). In sum, we have proved the nontrivial

THEOREM 1.1 (i) There are concrete systems. (ii) Every thing is a component of at least one system.

Surely the identification and modeling of a concrete system may be an extremely difficult task. Thus it is not always clear what the composition, hence also the environment, of a system is, particularly if it is strongly coupled to other systems – as is the case with the economic and the political systems of a society. However, this is a scientific problem not an ontological one.

Note that actions and the corresponding connections have been defined for things not for properties. The latter can be *interdependent* but not *interacting*. That is, the common phrase 'Properties P and Q interact' should be understood either as "Properties P and Q (of a given thing) are interdependent", or "Things with property P interact with things with property Q".

Connections can be permanent or temporary, static or dynamic. In the latter case they are often called *flows* – of energy, as in heat transfer, of matter, as in migrations, or of fields, as in a television network. If a physical flow happens to carry information, the connection is called *informational* and the entire system an 'information system'. However, the physical/informational distinction is one of emphasis, not a dichotomy, for every information flow rides on some energy flow. (See Appendix A, Sec. 1.4.)

Our definition of the environment of a system as the set of all things coupled with components of the system makes it clear that it is the *immediate* environment, not the *total* one – i.e. the set of all the things that are not parts of the system. Except in extragalactic astronomy and in cosmology, we are interested not in the transactions of a system with the rest of the universe but only in that portion of the world that exerts a significant influence on the thing of interest. This immediate environment or *milieu* is the cell in the case of chromosomes, the rest of the organism in the case on an organ, the ecosystem in the case of an organism, the solar system in the case of a biosphere, and so on. In other words, the immediate environment of a thing is the composition of its next supersystem. (More in Sec. 1.4.)

A system that neither acts on nor is acted upon by any other thing is said to be closed. In other words, we make

DEFINITION 1.3 Let σ be a system with environment $\mathscr{E}(\sigma, t)$. Then σ is *closed* at t iff $\mathscr{E}(\sigma, t) = \emptyset$ – otherwise σ is *open*.

Since every thing but the universe interacts with some other things, we infer

COROLLARY 1.1 The universe is the only system closed at all times.

This holds whether or not the universe turns out to be spatially infinite, for the universe may be defined as that thing which has a void environment (i.e. which is self-contained).

So much for the concept of total closure. We need also the notion of partial closure, or closure relative to a given property, since a system may be open in some respects and closed in others. (Thus all systems are gravitationally open, but some are electrically closed, others are closed to the exchange of matter, still others to cultural influences, and so on.) We make then

DEFINITION 1.4 Let P be a property of a system σ in an environment $\mathscr{E}(\sigma, t)$. Then σ is *open with respect to P at t* iff P is related, at t, to at least one property of things in $\mathscr{E}(\sigma, t)$ – otherwise σ is *closed* in the respect P.

Comparing this definition with the previous one we realize that a system is closed iff it is closed in every respect.

Finally some comments on the concept of structure. Our use of it is common in mathematics and in the social sciences. Thus a famous anthropologist: for the biochemist an organism "is a complexly integrated system of complex molecules. The system of relations by which these units are related is the organic structure. As the terms are here used the organism is *not* itself a structure; it is a collection of units (cells or molecules) arranged in a structure, i.e. in a set of relations; the organism *has* a structure" (Ratcliffe-Brown, 1935). Biologists use 'structure' sometimes in the above sense and at other times as a synonym for 'anatomic component'. In the latter case they run the risk of talking about the structure of a structure.

It is sometimes useful to distinguish a system's internal structure from its external one. The former is the subset of the total structure formed by the relations (in particular connections) among the system components. And the external structure is of course the complement of the internal structure to the total structure. Though distinct, the internal and the external structure are interdependent. Thus the internal structure of a molecule, far from being a permanent and intrinsic property of the molecule, depends critically upon its external structure – i.e. the interactions between the molecule and its milieu (e.g. the solvent).

Another distinction worth making is that between total structure and spatial structure, or set of spatial relations among the parts of a thing. (Spatial structure or configuration should not be mistaken for shape. The great majority of systems in the universe, i.e. the hydrogen and helium atoms, are shapeless. Nor do social systems have a shape although they have a spatial configuration since they are made up of living beings that stand in definite spatial relations to one another.) Every system has both a system structure (or bondage) and a spatial structure (or configuration). On the other hand aggregates or assemblages have spatial structures but no system structures.

For ease of reference we collect some of the above elucidations in

DEFINITION 1.5 Let σ be a concrete system with A-structure $\mathscr{S}_A(\sigma, t)$ at time t. Then

(i) the *internal A-structure* of σ at t is the subset of $\mathscr{S}_A(\sigma, t)$ composed of the relations among the A-parts of σ at t;

(ii) the *configuration* (or *spatial structure*) of σ at t is the subset of $\mathscr{S}_A(\sigma, t)$ composed of the spatial relations among the A-parts of σ at t.

1.4. *Subsystem*

A system component may or may not be a system itself. If it is we call it a 'subsystem'. More explicitly, we make

DEFINITION 1.6 Let σ be a system with composition $\mathscr{C}(\sigma, t)$, environment $\mathscr{E}(\sigma, t)$ and structure $\mathscr{S}(\sigma, t)$ at time t. Then a thing x is a *subsystem* of σ at t, or $x \prec \sigma$, iff

(i) x is a system at time t, and

(ii) $\mathscr{C}(x, t) \subseteq \mathscr{C}(\sigma, t)$ & $\mathscr{E}(x, t) \supseteq \mathscr{E}(\sigma, t)$ & $\mathscr{S}(x, t) \subseteq \mathscr{S}(\sigma, t)$.

By definition, the subsystem relation $\prec$ is an order relation, i.e. it is reflexive, asymmetric, and transitive. So, in particular, if $\sigma_1 \prec \sigma_2$ and $\sigma_2 \prec \sigma_3$, then $\sigma_1 \prec \sigma_3$. We shall make use of this property when defining the notion of a system of nested systems (Definition 1.7).

Example 1 Factories, hospitals and schools constitute subsystems of any modern society. On the other hand the persons composing them are not themselves social systems: they are biosystems. *Example 2* A foetus is a subsystem of its mother; it becomes a system in its own right after birth: before that it does not fall under any of the laws, natural or social, that hold for independent systems.

Systems of different kinds have different compositions or different structures. (A difference in composition induces a structural difference but not conversely, as shown by the existence of isomers, i.e. systems with the same composition but different structures.) However, all systems of the same genus seem to have the same overall structure or "general plan" –

pardon the anthropomorphism. For example, all atoms consist of nuclei surrounded by electrons, all solids are atomic or ionic lattices inhabited by wandering electrons, and even the overall structure of the skeleton and organs is the same for all vertebrates. (However, the precise characterization of the notion of overall structure is an open problem.)

Structures are often said to come superposed or nested like systems of Chinese boxes. Thus a polypeptide is said to have two structures, one primary or basic (the linear sequence of amino acids), the other secondary and consisting in the configuration of the entire coil. The helical configuration of the DNA molecule is an example of a secondary structure. In turn the secondary structure may determine a tertiary structure, e.g. the folding of the whole double strand into a regular configuration. See Figure 1.3.

In our view there is no such thing as a hierarchy of structures. (Etymologically 'hierarchy' means a set of sacred components ordered by a power or domination relation.) What we do have here is a system of nested systems, i.e. a collection of systems each of which is a subsystem of a larger system (or supersystem). And what molecular biologists call 'primary structure' is the structure of the innermost or core system, the secondary structure is the structure of the next supersystem, and so on. This notion is elucidated by

DEFINITION 1.7. Let σ be a system and call Σ the totality of systems, and

$$N_\sigma = \{\sigma_i \in \Sigma \mid \sigma \prec \sigma_i \,\&\, 1 \leq i \leq n\}$$

a collection of supersystems of σ partially ordered by the subsystem relation $\prec$. Then

(i) N_σ is a system of *nested* systems with core σ;

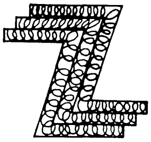


Fig. 1.3. An imaginary system of Chinese boxes or hierarchy of systems. The primary structure, i.e. the amino acid sequence, is not shown. The secondary structure is the helix, the tertiary the Z shape. And the quaternary structure is the way the individual Z's are assembled together, i.e. the double staircase.

(ii) the *primary structure* of σ is the structure of σ itself; the *secondary structure* of σ is the structure of the smallest supersystem of σ in N_σ, i.e. σ_1; in general, the *n-ary structure* of σ is the structure of σ_{n-1}.

1.5. *Level*

Talk of levels of organization (or complexity, integration, or evolution) and of a hierarchy of such has been rampant in science, particularly in biology, for the last half century or so. Unfortunately there is no consensus on the significance of the terms 'level' and 'hierarchy', which are used in a variety of ways and seldom if ever defined (Bunge, 1959b, 1959c). This fuzziness must be blamed not only on scientists but also on philosophers – on the inexact philosophers who despise clarity and on the exact ones who are not aware of the philosophical problems raised by scientific research. Let us attempt to remedy this situation by clarifying one concept of level, and the corresponding one of hierarchy, that are widely used in contemporary science.

The intuitive idea is simple: the things at any given level are composed of things belonging to preceding levels. Thus biospheres are composed of ecosystems, which are composed of populations, which are composed of organisms, which are composed of organs, which are composed of cells, which are composed of organelles, which are composed of molecules, which are composed of atoms, which are composed of so-called elementary particles. One way of exactifying this notion is by means of

DEFINITION 1.8 Let $L = \{L_i \mid 1 \leq i \leq n\}$ be a family of n nonempty sets of concrete things. Then

(i) one level *precedes* another iff all the things in the latter are composed of things in (some or all of) the latter. I.e. for any L_i and L_j in L,

$$L_i < L_j =_{df} (\forall x)[x \in L_j \Rightarrow (\exists y)\,(y \in L_i \;\&\; y \in \mathscr{C}(x))];$$

(ii) a thing *belongs to a given level* iff it is composed of things in (some or all of) the precedings levels. I.e. for any $L_i \in L$:

For any x in L_i: $x \in L_i =_{df} \mathscr{C}(x) \subset \bigcup_{k=1}^{i-1} L_k$;

(iii) $\mathscr{L} = \langle L, < \rangle$ is a *level structure.*

Note the following points. First, a level is not a thing but a set and therefore a concept, though not an idle one. Hence levels cannot act upon one another. In particular the higher levels cannot command or even obey the lower ones. All talk of interlevel action is elliptical or metaphorical, not literal. Second, the relation between levels is neither the part–whole relation nor the set inclusion relation but a *sui generis* relation definable in terms of the former. Third, there is nothing obscure about the notion of level precedence as long as one sticks to the above definition instead of construing '$L_i < L_j$' as "the L_i's are inferior to the L_j's" or in similar guise. Fourth, it is mistaken to call a level structure $\mathscr{L} = \langle L, < \rangle$ a *hierarchy*, because the level order $<$ is not a dominance relation (Bunge, 1973a). Fifth, our concept is so far static: we are not assuming anything about the origin or mode of composition of systems in terms of evolution.

1.6. *Systems Association*

Whether or not two things form a system, they can be assumed to associate (or add physically) to form a third thing. Thus thing a and thing b, no matter how distant and indifferent, may be assumed to form thing $c = a \dotplus b$. In other words, the set of things is closed under the operation $\dotplus$ of *association*, physical addition, or juxtaposition (Vol. 3, Ch. 1, Sec. 1).

Not so with systems: two systems may or may not associate to form a third. Thus two molecules may not combine to form a system, and two social systems may not merge to form a third. In general the physical addition or association of two things will be a thing but not a system: *systemicity is not conserved.* The environment, the structure and perhaps even the composition of the resulting thing are different from the mere union of the partial compositions, environments, and structures. See Figure 1.4. In short, the set of all systems has no algebraic structure – not even the rather modest one of a semigroup. But, of course, since systems are things, they do comply with the algebra of things. In particular they associate to form further things.

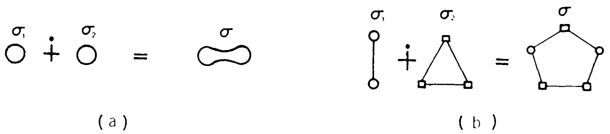


Fig. 1.4. (*a*) Before fusion, $\mathscr{E}(\sigma_1) = \{\sigma_2\}$, $\mathscr{E}(\sigma_2) = \{\sigma_1\}$. After fusion, $\mathscr{E}(\sigma_1 + \sigma_2) = \varnothing$. (*b*) Before merger, $\mathscr{S}(\sigma_1) = \{\text{Linear link}\}$ and $\mathscr{S}(\sigma_2) = \{\text{Triangular link}\}$. After merger, $\mathscr{S}(\sigma_1 + \sigma_2) = \{\text{Pentagonal link}\}$.

1.7. *Other Kinds of System: Property and Functional*

We are interested not only in concrete systems but also in property systems, or sets of interrelated properties, as well as in functional systems, or sets of coupled processes. For example, most of the properties of a thing that is either simple (basic) or a system hang together, i.e. a change in one of them is accompanied by changes in others. As a consequence most of the changes occurring in either a simple thing or a system are coupled, so that if one of them starts or stops, others change. (The cautious prefix 'most' is intended to exclude superficial properties, such as position and color, which can often change considerably, within bounds, without dragging changes in other properties.) Although every thing other than an aggregate or conglomerate has properties and undergoes processes that constitute systems, property systems and functional systems are particularly conspicuous among organisms. In particular the mental abilities of an animal form a system.

We shall use the following conventions:

DEFINITION 1.9 Let $p(x)$ be the set of properties of a thing x, and $\pi(x)$ the set of processes occurring in x. Then

(i) the subset $p_0(x) \subset p(x)$ is a *property system* of x iff every property in $p_0(x)$ is lawfully related to at least one other property in $p_0(x)$;

(ii) the subset $\pi_0(x) \subset \pi(x)$ is a *functional system* of x iff every process in $\pi_0(x)$ is lawfully related to at least one other process in $\pi_0(x)$.

Because in a system all properties and processes are lawfully interrelated we conclude that, for every x, x is a concrete system iff $p(x)$ is a property system or $\pi(x)$ a functional system.

1.8. *Concluding Remarks*

The literature on systems is vast, rapidly growing, and somewhat bewildering. (Cf. Klir and Rogers, 1977.) However, the field is still immature and its reputation is jeopardized by a fringe of charlatans. Suffice it to mention three indicators of immaturity.

Firstly, the very definition of the concept of a system is still in doubt, so that many a paper starts by spending time defining or redefining the concept. Yet, so much effort spent on definitions has yielded only three which are as popular as they are incorrect. According to the first definition, a system is a set of interrelated elements – which is fine for conceptual systems but not for concrete ones since sets, no matter how structured, are sets, hence concepts not things. The second definition equates a system with a black box equipped with inputs and outputs, which is fine in a few cases but useless when the internal structure of the system is relevant. And the third widely used definition is a generalization of the preceding, namely this: a system is a binary relation – again a conceptual object.

Secondly, some writers claim that everything imaginable is a system, and that a general theory of systems should deal with every possible thing (without thereby becoming part of philosophy) and every possible problem, theoretical or practical, concerning the behavior of systems of all kinds. Some have even asserted that such a theory should cover not only concrete systems but also conceptual ones, so that it would be a thoroughly unified science of everything.

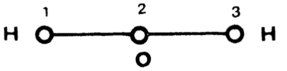
Thirdly, some enthusiasts of general theories of systems have seen in these a vindication of holistic philosophies, hence a condemnation of the analytic method characteristic of science. However, most of those who approve of general systems theories for their alleged holistic virtues either misuse the term 'holistic' to designate "systemic", or are interested in instant wisdom rather than painstaking scientific or philosophical research.

Such confusions and wild claims, which linger because of insufficient foundational research in the field of systemics, have elicited some entirely negative reactions to it (e.g. Berlinski, 1976). While there is some legitimacy in such reactions, there is no denying that systemics abounds in good theories – such as automata theory and general lagrangian dynamics – serviceable in a number of fields, and that it provides an inspiring framework for posing problems and building models. Rather than throw out the baby together with the bath water we ought to change the latter once in a while.

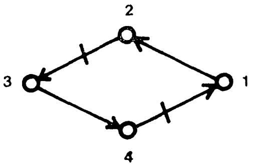
2. SYSTEM REPRESENTATIONS

2.1. *Coupling Graphs and Matrices*

We shall presently review two standard and equivalent ways of representing a system with a denumerable composition, be it a molecule or an industrial plant. They are the graph and the matrix representations. (See Klir and Valach (1967).) The following examples show how to proceed.

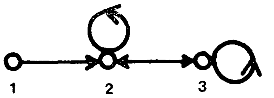


$$\sigma_1 = \begin{Vmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{Vmatrix}$$



$$\sigma_2 = \begin{Vmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \\ -1 & 0 & 0 & 0 \end{Vmatrix}$$

The arrows indicate excitation, the crossed arrows inhibition.



$$\sigma_3 = \begin{Vmatrix} 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 1 & 1 \end{Vmatrix}$$

The loops indicate self action or feedback.

We generalize the foregoing into the following semantic assumptions. Let σ be a system with m components and n different kinds of connection among them (e.g. mechanical, chemical, informational, social, etc.). Then σ is representable by

(i) a set of n directed graphs over the composition of σ, one for each kind of connection, with a total of m nodes (vertices), in such a way that (*a*) the nodes represent the components and (*b*) the edges represent the connections; or

(ii) a set of n $m \times m$ matrices ${}^pM$, where $1 \leq p \leq n$, such that (*a*) the matrix element ${}^pM_{rs}$ of the pth matrix represents the strength of the action of component r, in the pth respect, upon component s, and (*b*) the matrix element ${}^pM_{rr}$ stands for the action of kind p of the rth component on itself (feedback).

The off-diagonal elements ${}^pM_{rs}$, with $r \neq s$, represent connections other than self connections. There are $m^2 - m = m(m - 1)$ such elements per matrix, and a total of $nm(m - 1)$ per system with n different kinds of connection. This number is called the *coupling capacity* of the system.

So far we have represented the composition and the internal structure of a system while neglecting its environment, hence its external structure. An open system, i.e. one connected with its environment, can be represented as follows. Instead of building an $m \times m$ matrix for an m component system, as directed by the previous semantic postulate, we form an $(m + 1) \times (m + 1)$ matrix for each kind of connection, letting 0 stand for the environment *en bloc*. Any system component r for which $M_{0r} \neq 0$ is an input or receiver component, whereas s is an output or donor component of the system if $M_{s0} \neq 0$. For example, a two component open system with a single kind of connection can be represented by the matrix

$$M = \begin{Vmatrix} 0 & M_{01} & M_{02} \\ M_{10} & M_{11} & M_{12} \\ M_{20} & M_{21} & M_{22} \end{Vmatrix}.$$

The elements M_{01} and M_{02} are the inputs (to the first and second components respectively) and the entries M_{10} and M_{20} the outputs (of the first and second components respectively). The other entries represent the internal (or internuncial) connections among the system's components.

We generalize the preceding into the following semantic assumption. Let σ be a system with m components and n different kinds of connections among them. Moreover, let the environment of σ be construed as a single entity labelled 0. Then σ is representable by n $(m + 1) \times (m + 1)$ matrices ${}^pM$, where $1 \leq p \leq n$, such that

(i) the *internal connectivity* of σ in the pth respect is representable by the matrix obtained from ${}^pM$ by striking off the M_{r0} and M_{0s} elements;

(ii) the *input* to σ in the respect p is represented by the row of input entries of ${}^pM$, i.e.

$${}^p\mathscr{I}(\sigma) = \left\| {}^pM_{01} \quad {}^pM_{02} \ldots {}^pM_{0m} \right\|;$$

(iii) the *output* of σ in the respect p is represented by the column of output entries of ${}^pM$, i.e.

$${}^p\mathscr{O}(\sigma) = \left\| {}^pM_{10} \quad {}^pM_{20} \ldots {}^pM_{m0} \right\|^t,$$

where t designates the transposition operation (conversion of row matrix into column matrix);

(iv) the *behavior* (or *observable performance*) of σ in the respect p is the ordered pair

$${}^p\beta(\sigma) = \langle {}^p\mathscr{I}(\sigma), {}^p\mathscr{O}(\sigma) \rangle;$$

(v) the (total) *behavior* of σ is the set of its partial behaviors:

$$\beta(\sigma) = \{{}^p\beta(\sigma) \mid 1 \leq p \leq n\}.$$

Example In the simplest case, of a two component system interacting with its environment in a single way, we have

$$\mathscr{I}(\sigma) = \|M_{01} \quad M_{02}\|, \qquad \mathscr{O}(\sigma) = \begin{Vmatrix} M_{10} \\ M_{20} \end{Vmatrix}.$$

In the absence of any data or hypotheses concerning the internal structure (i.e. the full coupling matrix) of such a system, we must restrict our attention to its behavior. The best we can do is to guess that the latter is linear, i.e. that there exists a matrix T transforming inputs into outputs: $\mathscr{O} = T\,\mathscr{I}^t$, where $\mathscr{I}^t$ is the transpose of $\mathscr{I}$. We set then

$$T = \begin{Vmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{Vmatrix}$$

with unknown T_{ij}'s and perform the indicated operations:

$$T\,\mathscr{I}^t = \begin{Vmatrix} T_{11} & T_{12} \\ T_{21} & T_{22} \end{Vmatrix} \cdot \begin{Vmatrix} M_{01} \\ M_{02} \end{Vmatrix} = \begin{Vmatrix} T_{11}M_{01} + T_{12}M_{02} \\ T_{21}M_{01} + T_{22}M_{02} \end{Vmatrix} = \begin{Vmatrix} M_{10} \\ M_{20} \end{Vmatrix},$$

thus obtaining the algebraic system

$$\begin{aligned} T_{11}M_{01} + T_{12}M_{02} &= M_{10} \\ T_{21}M_{01} + T_{22}M_{22} &= M_{20}. \end{aligned}$$

This system of equations has no unique solution when only the behavior of the concrete system (i.e. the M_{0i}'s and the M_{j0}'s) are given, for in this case there are only two conditions (equations) for four unknowns (the T_{ij}'s). Even hitting upon a solution by trial and error won't advance us a single step in the process of finding out the structure of the system, i.e. the complete coupling matrix M. The only procedure that could bring success is to guess and try out alternative assumptions about the system's structure and check whether they do yield the observed (or conjectured) behavior. That is, the way to theoretical knowedge is not from behavior to inferred structure but from hypothesized structure to behavior. This shows that behaviorism, phenomenalism, and inductivism are incapable, not just unwilling, to explain behavior.

Obviously neither the graph nor the matrix representation of a system suffices for all purposes. It represents only the composition, structure, and environment of a system with neglect of its dynamics. A more complete representation can only be obtained by setting up a full fledged dynamical theory incorporating and expanding the information contained in the graph or the matrix representation. We turn next to the common core of such dynamical representations, namely the state space representation. (See the Appendices for a number of particular, yet also cross-disciplinary, mathematical models of systems.)

2.2. *The State Space Representation*

Every system of a given kind K has a finite number n of general properties, such as age, number of components, connectivity among them, inputs, and outputs. And each such general property is representable by a function $F_i\colon A \to V_i$, where $1 \leq i \leq n$. Collecting all such property-representing functions into a single ordered n-tuple or list

$$\mathbb{F} = \langle F_1, F_2, \ldots, F_n \rangle : A \longrightarrow V_1 \times V_2 \times \ldots \times V_n,$$

we form the *state function* of systems of the given kind. Just as $\mathbb{F}$ represents the totality of general properties of the K's, so each value $\mathbb{F}(a) = \langle F_1(a), F_2(a), \ldots, F_n(a) \rangle$ represents the totality of individual properties of a particular system, such as its age and composition at a given time.

The domain A of the state function $\mathbb{F}$ of systems of kind K is the cartesian product of certain sets, such as K, the family 2^E of sets of environmental items with which the members of K are coupled, the set F of reference frames, the set T of time instants, and so on. (2^E is the power set of the set E of environmental things, so the environment e of a particular system is a member of that family, i.e. $e \in 2^E$.) And the codomain V_i of the ith component F_i of the state function is usually taken to be some subset of the real line $\mathbb{R}$. (If a property is represented by a complex valued function each component of the latter counts as a component of $\mathbb{F}$.) In short,

$$\mathbb{F} : K \times 2^E \times F \times T \times \cdots \longrightarrow \mathbb{R}^n.$$

The value $\mathbb{F}(k, e, f, t, \ldots) = \langle a, b, \ldots, n \rangle \in \mathbb{R}^n$ of the state function of system $k \in K$ interacting with environmental items $e \in 2^E$, relative to reference frame $f \in F$ at time $t \in T$, is the *state* of k at t. The collection of all such possible states, which is a subset of $\mathbb{R}^n$, is the (conceivable) *state space* of systems of kind K, or $S(K)$ for short. However, since the components of $\mathbb{F}$ are lawfully interrelated, and thus mutually restricted, not every n-tuple of real numbers represents a really (or nomologically) possible state of a system. I.e. the *lawful state space* of systems of kind K, or $S_L(K)$ for short, is a proper subset of the conceivable state space $S(K)$.

In short, every really possible state of a K is a point in some region $S_{\mathbb{L}}(K)$ of the cartesian space $\mathbb{R}^n$. See Figure 1.5.

Example 1 In the elementary kinetic theory of gases, the state function is the triple consisting of the pressure, volume, and temperature functions. The corresponding state space is a cube contained in $(\mathbb{R}^+)^3$. *Example 2* In Hamiltonian dynamics the state (or phase) vector is $\langle q(k, f, t), p(k, f, t)\rangle$, where q is the canonical coordinate and p the corresponding momentum – neither of which need be mechanical properties. *Example 3* In chemical kinetics the instantaneous state of a chemical system is described by the values of the partial concentrations of reactants and reaction products. Therefore the state space of the system is a hypercube contained in $(\mathbb{R}^+)^n$, where n is the number of system components (reactants, catalyzers, and products). If there is diffusion, further axes must be added to the state space, in particular temperature and position coordinates. *Example 4* In the genetics of populations three commonly used state variables are the size of a population, the probability (incorrectly called "frequency') of some particular gene or gene constellation, and the latter's adaptive value. Hence for a system composed of two interacting populations, A and B, the state space is the region of $\mathbb{R}^6$ spanned by the sextuples $\langle N_A(t), N_B(t), P_A(t), P_B(t), v_A(t), v_B(t)\rangle$ in the course of time.

The concept of state space can be used to clarify that of system. The state space of an aggregate or conglomerate of non-interacting things is uniquely determined by the partial states spaces. Moreover, since the

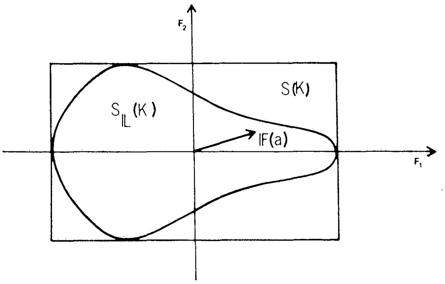


Fig. 1.5. The lawful state space $S_{\mathbb{L}}(K)$ of systems of kind K is a subset of the cartesian product of the ranges of the components of the state function. Only two such components, F_1 and F_2, are depicted here. $\mathbb{F}(a) = \langle F_1(a), F_2(a)\rangle$ is a (really) possible state of a particular system of the given kind. As time 'goes by', the tip of $\mathbb{F}(a)$ moves about within $S_{\mathbb{L}}(K)$.

contributions of the latter are all on the same footing, we may take the total state space to equal the union of the partial state spaces. In particular, let $S_{\mathbb{L}}(K)$ and $S_{\mathbb{L}}(M)$ be the lawful states spaces of things of kinds K and M respectively. Then the state space of the association $k + m$ of two non-interacting things of kinds K and M respectively, relative to the same reference frame, is $S_{\mathbb{L}}(K) \cup S_{\mathbb{L}}(M)$. Not so in the case of a system: here the state of every component is determined, at least partly, by the states other system components are in, so that the total state space is no longer the union of the partial state spaces. Thus in Example 4 above the state space of the two component system must be constructed *ab initio* rather than on the sole basis of the state spaces for the individual biopopulations. In sum, a thing is an *aggregate* if and only if its state space equals the union of the state spaces of its components – otherwise it is a (concrete) *system.* (Cf. Bunge, 1977a, 1977b.)

Every event occurs in or to some concrete thing and it consists in a change of state of the thing – the change being merely quantitative, as in the case of motion, or qualitative as well, as in the case of the coming into being or the metamorphosis of a thing. A light flash, the dissociation of a molecule, a storm, the growth of a bud, the learning of a trick, and the fall of a government are events – or rather processes, for they are complex and therefore analyzable into further events. Being changes in the states of things, events and processes are representable as trajectories in the state spaces of changing things. (An unchanging thing, if there were one, would have a state space consisting of a single point.) Different trajectories in a state space may have the same end points. That is, there are cases in which one and the same net change can be effected along alternative routes. See Figure 1.6.

The functions g and g' occurring in Figure 1.6 are not supposed to be arbitrary: they must be lawful if we are to allow only lawful events and discard lawless ones, i.e. miracles. In other words, the g occurring in the event representation $e = \langle s, s', g\rangle$ must be compatible with the laws of the system(s) concerned. Equivalently: a *lawful event or process* occurring in a system of kind K, with end points s and s', is representable by a triple $\langle s, s', g\rangle$, where $g : S_{\mathbb{L}}(K) \rightarrow S_{\mathbb{L}}(K)$ is compatible with the laws of the K's. If we disregard the intermediate states between the end points of the processes, we are left with arrows or ordered pairs $\langle s, s'\rangle \in S_{\mathbb{L}}(K) \times S_{\mathbb{L}}(K)$. The collection of all such pairs of states, i.e. the set of all net events (for a given g) constitutes the *event space* of the K's (for g). Symbol:

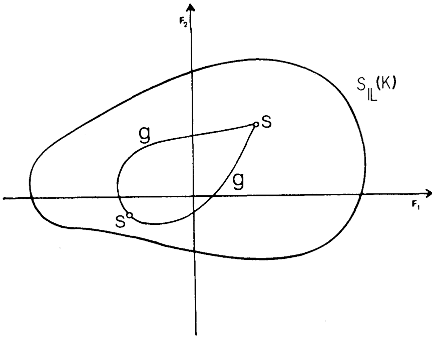


Fig. 1.6. Two different processes resulting in the same net change. The net change from state s to state s' can be represented as the ordered pair (or arrow) $\langle s, s' \rangle$. Since the change along curve g may be distinct from the change along curve $g' \neq g$, we must represent the full events (or processes) by $\langle s, s', g \rangle$ and $\langle s, s', g' \rangle$ respectively.

$E_{\mathbb{L}}(K) \subseteq S_{\mathbb{L}}(K) \times S_{\mathbb{L}}(K)$. In general the inclusion is proper: not all state transitions are lawful.

Since we shall make ample use of the state space representation in this work, we may as well compress the preceding into the following semantic assumption. For each kind K of system possessing n properties there is a property-representing function $\mathbb{F}: A \to V_1 \times V_2 \times \ldots \times V_n$ with n components, called a *state function* for systems of the kind. Furthermore

(i) the *totality of general properties* of systems of kind K is representable by the set of all the components (or coordinates) of $\mathbb{F}$, i.e. $p(K) = \{F_i \mid 1 \leq i \leq n\}$;

(ii) every *particular property* of a system of kind K is representable by a value of a component of $\mathbb{F}$, i.e. by $F_i(a)$ for $a \in A$ and some $1 \leq i \leq n$;

(iii) the *state* of systems of kind K at $a \in A$ is representable by the value of $\mathbb{F}$ at a, i.e. $s = \mathbb{F}(a) = \langle F_1(a), F_2(a), \ldots, F_n(a) \rangle$;

(iv) the collection of all such states of things of kind K, i.e. the range of $\mathbb{F}$, is the *lawful state space* of the K's, or $S_{\mathbb{L}}(K)$ for short;

(v) every *event* occurring in a system of kind K is representable by an ordered triple $\langle s, s', g \rangle$, where $s, s' \in S_{\mathbb{L}}(K)$ and g is a lawful map of $S_{\mathbb{L}}(K)$ into itself;

(vi) the collection of all really possible (lawful) events occurring in systems of kind K is the *event space* of K, or $E_{\mathbb{L}}(K)$ for short;

(vii) for a system in a given environment, and relative to a given reference frame, the state function takes often the form of a time-dependent function $\mathbb{F}: T \to \mathbb{R}^n$, where $T \subseteq \mathbb{R}$ is the set of instants relative to the given frame;

(viii) If $\mathbb{F}: T \to \mathbb{R}^n$, then the totality of *processes* occurring in a system x of kind K during the time interval $\tau \subseteq T$ is representable by the set of states x is in during τ:

$$\pi(x, \tau) = \{\mathbb{F}(t) \mid t \in \tau\};$$

(ix) the *history* of a system x of kind K representable by a state function $\mathbb{F} : T \to \mathbb{R}^n$, during the interval $\tau \subseteq T$, is representable by the trajectory

$$h(x) = \{\langle t, \mathbb{F}(t)\rangle \mid t \in \tau\};$$

(x) the *total action* (or *effect*) of a thing x on a thing y equals the difference between the forced trajectory and the free trajectory of the patient y:

$$A(x, y) = h(y \mid x) \cap \overline{h(y)}.$$

A detailed treatment of these concepts is given elsewhere (Vol. 3, Ch. 5). We shall presently use them to advance a handful of general principles concerning systems.

3. BASIC ASSUMPTIONS

3.1. *Structural Matters*

So far we have made only a few definitions and semantic assumptions but no substantive hypotheses on the nature of systems. (Theorem 1.1 on the existence of systems and the nonexistence of stray things, as well as Corollary 1.1 on the openness of systems, followed from our definition of a concrete system in conjunction with certain general postulates, about the nature of things, laid down in Volume 3.) We shall presently wager a handful of basic assumptions concerning systems of all kinds, the first being certain postulates of a structural kind. Since these assumptions will concern the transactions of a system with its environment, we may as well define the concepts of input and of output in more general terms than we did in Sec. 2.1. We start then with

DEFINITION 1.10 Let σ be a system with an (immediate) environment $\mathscr{E}(\sigma)$. Then

(i) the totality of *inputs* of σ is the set of all the environmental actions on σ:

$$U(\sigma) = \bigcup_{x \in \mathscr{E}(\sigma)} A(x, \sigma);$$

(ii) the totality of *outputs* of σ is the set of all the actions of the system on its environment:

$$V(\sigma) = \bigcup_{y \in \mathscr{E}(\sigma)} A(\sigma, y);$$

(iii) the *activity of the environment* of σ is

$$E(\sigma) = \bigcup_{x, y \in \mathscr{E}(\sigma)} A(x, y) \cup U(\sigma) \cup V(\sigma).$$

Our first hypothesis is that all systems receive inputs and are selective, i.e. accept only a (small) subset of the totality of environmental actions impinging on them. More precisely, we lay down

POSTULATE 1.1. Let σ be a system with total input $U(\sigma)$. Then
(i) $U(\sigma) \neq \emptyset$;
(ii) $U(\sigma) \subset E(\sigma)$ or, equivalently, the (in-selection) function

$$\mu: U(\sigma) \longrightarrow E(\sigma)$$

is the inclusion (or embedding) map of $U(\sigma)$ into $E(\sigma)$.

Example Talking to plants is ineffectual – except insofar as it nourishes them with water and carbon dioxyde.

A second, equally pervasive, feature of concrete systems is that they react on their environment, i.e. that their output is never nil. (The so-called machines without output, studied in automata theory, are of course fictions.) Moreover, in every system there is spontaneous activity, i.e. not elicited by any inputs. Thus we make

POSTULATE 1.2. Let $V(\sigma)$ be the total output of a system σ. Then
(i) $V(\sigma) \neq \emptyset$;
(ii) the (out-selection) function

$$\nu: V(\sigma) \longrightarrow E(\sigma)$$

assigns each system output an environmental action but not conversely.

Example In every neuron there is spontaneous activity that is superposed on the activity elicited by afferent stimulation.

The previous hypotheses are typical metaphysical principles in as much as they can be confirmed but not refuted, for they hinge on an item that is only partially knowable, namely the set E of environmental actions. Any evidence unfavorable to our postulates may be blamed on our ignorance of most of E.

The in-selection function μ and the out-selection function ν are joined in

DEFINITION 1.11 The function f that composes with the out-selection function ν of a system to yield its in-selection function μ, i.e. such that $\mu = f \circ \nu$, is called the *transfer* (or *transducer*) function f: $U(\sigma) \rightarrow V(\sigma)$ of σ:

$$\begin{tikzcd}
U(\sigma) \arrow[r, "\mu"] \arrow[dr, "f"'] & E(\sigma) \\
& V(\sigma) \arrow[u, "\nu"']
\end{tikzcd}$$

Example The retina transforms (or maps or codes) light stimuli into nervous signals.

We close this subsection with a batch of general principles that will be stated in an informal way:

POSTULATE 1.3 Let σ be an arbitrary system other than the universe. Then

(i) every input to σ is an output of some other system (i.e. there are no inputs out of the blue);

(ii) σ receives inputs of several kinds (i.e. at some time or other every one of the components of the state function of σ is bound to be affected by environmental changes);

(iii) for every action upon σ there is a threshold below which σ does not respond;

(iv) the total input of σ has a nonvanishing random component;

(v) there is a delay, however small, between every input and the corresponding output if any.

So much for our general structural assumptions. Let us now look at systems from an evolutionary perspective.

3.2. *Assembly and Emergence*

Any process whereby a system is formed out of is components is called 'assembly'; if the process is spontaenous, it is called 'self-assembly'. An assembly process can occur either in a single step or, more likely, in a number of steps: see Figure 1.7. We can express the idea formally with the help of the concept of bondage, or set of bonds among the components of a system, introduced in Sec. 1.2. Indeed we make

DEFINITION 1.12 Let x be a concrete thing composed initially of uncoupled parts (possibly systems themselves), i.e. such that $\mathbb{B}(x, t) = \varnothing$. Then

(i) *x assembles into y* at time $t' > t$ iff y is a system with the same composition as x but a nonempty bondage, i.e.

$$\mathscr{C}(y, t') = \mathscr{C}(x, t) \,\&\, \mathbb{B}(y, t') \neq \varnothing;$$

(ii) the assembly process is one of *self-assembly* iff the aggregate x turns by itself [i.e. naturally rather than artificially] into the system y;

(iii) the self-assembly process is one of *self-organization* iff the resulting system is composed of subsystems that were not in existence prior to the onset of the process.

Assembly processes can be natural or artificial, and those of the latter kind in turn experimental (or laboratory) or industrial. Artificial assembly processes are of course man-guided. However, there are degrees of control. It is one thing to assemble a machine from its parts and another to assemble a molecule from its precursors. In most cases the latter process

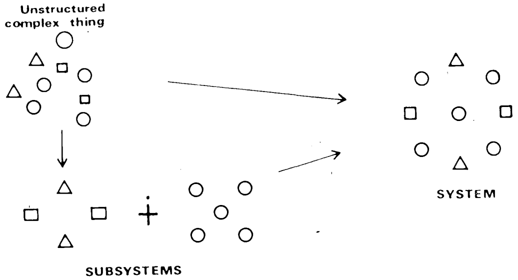


Fig. 1.7. Assembly of a system out of previously unconnected units, either directly or step-wise.

proceeds on its own – i.e. by virtue of its inner dynamics – once suitable reactants and physical conditions have been supplied. For example, proteins and even ribosome units will self-assemble *in vitro* in a matter of minutes if suitable precursors and physical conditions are provided. In this case man only lends a hand to nature, by reproducing conditions that have or could have occurred spontaneously.

All chemical and biochemical syntheses are of course assembly processes and moreover processes accompanied by the emergence of new properties. However, self-assembly occurs on all levels in a variety of manners. Perhaps the most conspicuous of all self-assembly process is the aggregation or clumping of atoms and cosmic dust brought about by gravitation. This is how planets and other celestial bodies are believed to have formed. Moreover, far from always leading to disorganized lumps, gravitational attraction can give rise to complex clustering on all astronomical scales, in particular star clusters, galaxies, and galaxy clusters. (See, e.g. de Vaucouleurs, (1970).) Similar examples of self-assembly by condensation of units of the same kind are the formation of molecular complexes (such as the haemocyanine polymolecule), of polymers such as the polypeptides and of crystals out of solutions. (See Calvin (1969); Eigen (1971); Lehninger (1975).) Needless to say, self-assembly, and in particular self-organization, occurs also on the social level: witness the formation of families, bands, communities, and social organizations of various kinds. In sum self-assembly and self-organization are not exclusive to life. What is peculiar to biotic self-organization is that it happens to ensue in living systems rather than in biochemical systems or in systems of some other kind. (In other words, it is an emergence mechanism leading to a new level of organization.) We shall assume that all systems short of the entire world have formed by assembly:

POSTULATE 1.4 All systems, except the universe, originate by assembly.

Remark 1 Natural systems originate by self-assembly, and artificial systems by artificial or man-made assembly. *Remark 2* An exception is made for the universe because (*a*) in naturalistic cosmologies the universe has neither an origin nor an end, and (*b*) in religious cosmologies it makes no sense to say that the universe originates by assembly, as this presupposes the prior existence of its components. *Remark 3* Surely a system may be formed by the breakdown of some supersystem. However, this is no counterexample to our postulate, which requires only that the original supersystem was the result of some assembly process. *Remark 4* Our axiom is far from obvious and it could not have been formulated before systems-theoretic thinking became pervasive, if only because the general concept of a system had not been elucidated until recently. *Remark 5* Our axiom is of particular relevance to the problem of the origin of biomolecules and biosystems. Until recently both were assumed either to have been created by divine *fiat* or to have existed from all eternity. Darwin's *Origin of Species* appeared only in 1859, and scientific research into the evolution of molecules started only one century later.

The components of a self-assembled system are called its *precursors*, an apt name suggesting that the system had not always been around but emerged from pre-existing things. Interestingly enough, (*a*) the precursors of a system are not blended but keep their individuality to some extent, and yet (*b*) they give rise to a thing possessing emergent properties. For example, the two hydrogen atoms that combine into a hydrogen molecule are distinct components of the latter, yet the spectrum of the molecule is radically different from that of its components. The first characteristic is formalized by saying that the composition of the self-assembled system equals the set of parts of its ancestors – i.e. the set of its precursors. And the notion of emergence can be elucidated as follows.

Call x a thing and $t \in T$ an instant of time, and introduce a function p that assigns the ordered couple $\langle x, t\rangle$ the set $p(x, t)$ of all the properties of x at t. That is, p is a function $p\colon \Theta \times T \to 2^{\mathbb{P}}$, where Θ is the set of all things, T the set of all instants, and $2^{\mathbb{P}}$ the family of subsets of the set $\mathbb{P}$ of all general properties of things. A change in thing x can be viewed as a certain change of state of x. Since x is held fixed throughout that change of state, we can use the simpler function

$$p_x\colon T \to 2^{\mathbb{P}} \quad \text{such that } p_x(t) = p(x, t).$$

In short, $p_x(t)$ is the collection of properties of thing x at time t. (For details on p_x, see Vol. 3, Ch. 2.)

Now let t and t' be two distinct instants, such that t precedes t'. The corresponding values of p_x are of course $p_x(t)$ and $p_x(t')$. If these two sets of properties of x are the same, then the thing has not changed qualitatively. If they are different, then the thing has gained or lost some properties. If the latter is the case, the newly gained properties will be said to be *emergent* relative to the given thing, even though they may be possessed by other things as well. In sum, we make

DEFINITION 1.13 Let x be a thing with properties $p_x(t)$ at time t, and properties $p_x(t')$ at a later time $t' > t$. Then

(i) the *total qualitative novelty* occurring in x during the interval $[t, t']$ is the symmetric difference

$$n_x(t, t') = p_x(t) \triangle \bigcup_{t<\tau\leq t'} p_x(\tau);$$

(ii) the *emergent properties* that appear in x during the interval $[t, t']$ are those in

$$e_x(t, t') = \bigcup_{t<\tau\leq t'} p_x(\tau) - p_x(t).$$

Example 1 Every nuclear and every chemical reaction results in the emergence of things endowed with emergent properties. *Example 2* The breakdown (dismantling) of a system, and the substitution of some of its components, are emergence processes.

We make bold and generalize:

POSTULATE 1.5 Every assembly process is accompanied by the emergence of some properties and the loss of others. I.e. let the parts of a thing x self-assemble into a system during the interval $[t, t']$. Then the system lacks some of the properties of its precursors – i.e. $p_x(t) - p_x(t') \neq \varnothing$–but on the other hand it possesses some new properties – i.e. $p_x(t') - p_x(t) \neq \varnothing$.

Thus far we have been concerned with qualitative novelty in a particular thing regardless of whether the given emergent properties are possessed by any other things. We now elucidate the concept of emergence for the first time, or absolute emergence:

DEFINITION 1.14 The *absolutely emergent properties* (or firsts) appearing in a thing x during the lapse $[t, t']$ are those in

$$e_x^a(t, t') = e_x(t, t') - \bigcup_{y\in\Theta} \bigcup_{\tau\in T} p_y(\tau) \quad \text{with } y \neq x \quad \text{and} \quad \tau \leq t'.$$

3.3. *Selection*

New systems are being formed all the time, but not all of them are viable in the environment in which they emerge. In fact many are unfit, hence ephemeral, and this either because they are internally unstable or because they cannot cope with environmental aggression. In the latter case we have to do with selection. In other words, the environment selects the fittest systems, be they molecules or men. We can formulate this idea with more precision in the following way.

We assume that every environment performs a selective or filtering action on any population of systems of some kind. This selective action consists in the shrinking of the original population S to some subset A of S, namely the collection of viable or adapted members of S. A permissive environment is such that A is almost as big as the original S, whereas a harsh environment will reduce A to a very small, possibly empty, subset A of S. We condense these ideas into a definition and a postulate.

DEFINITION 1.15 Let S be a set of systems of a given kind K, assembled during some time interval in a given environment $E = \mathcal{E}(\sigma)$ common to all $\sigma \in S$. Further, call $i_E\colon S \to A_E$ the inclusion function from S into A_E, where $A_E \subseteq S$. (I.e. $i_E(x) = x$ for any x in A_E.) Then

(i) the environment E exerts the *selectivw action*

$$i_E\colon S \longrightarrow A_E$$

on the population S iff, during the next time interval, only the members of A_E remain in S;

(ii) A_E is the set of systems of kind K *selected by* (or *adapted to*) the environment E, and $\overline{A_E} = S - A_E$ the collection of systems of the same kind *eliminated by* (or *ill adapted to*) E, during the given time interval;

(iii) the *selection pressure* exerted by E on S is the number

$$p(S, E) = \frac{|\overline{A_E}|}{|S|} = 1 - \frac{|A_E|}{|S|},$$

where '$|X|$' designates the numerosity of the set X.

Clearly, the p values range between 0 (maximal fitness) and 1 (maximal unfitness).

Our promised hypothesis reads thus:

POSTULATE 1.6 All systems are subject to environmental selection (or to a nonvanishing selection presssure). I.e. for every set S of any kind K and every environment E common to the members of S, there is a selective action function $i_E\colon S \to A_E$ with $A_E \subset S$.

An immediate consequence of this axiom together with Postulate 1.4 is

COROLLARY 1.2 All self-assembled systems are subject to environmental selection (or to a nonvanishing selection pressure).

Different environments may exert different selective pressures on one and the same system population. (This is why for every S we have as many inclusion maps i_E as possible environments of the S's.) For example, two different habitats, or two different seasons in the same habitat, may exert different selective actions upon one and the same population of organisms. And such actions compose in accordance with

THEOREM 1.2 Let E and E' be two different consecutive environments of members of a population S of systems of some kind, and let i_E and $i_{E'}$ be their respective selective actions (during the corresponding consecutive periods). Then the resulting selective action is the composition of the two partial selective actions, i.e.

$$i_{EE'} = i_{E'} \circ i_E,$$

and the corresponding selection pressure equals

$$p(S, EE') = p(S, E) \cdot p(A_E, E') = \left(1 - \frac{|A_E|}{|S|}\right) \cdot \left(1 - \frac{|A_{E'}|}{|A_E|}\right).$$

It should be obvious that the selective actions of two environments do not commute in general. So much so that, if the first environment is totally hostile, there is nothing left for the second to select from. The last environment always has the last word. See Figure 1.8.

It is often maintained that the environment is not creative, for all it is capable of is weeding out the maladapted. This is false. Firstly, the environment of every system includes further systems, some of which are capable of acquiring new properties. Secondly, every new system is as-

$$\begin{tikzcd}
S \arrow[r, "i_E"] \arrow[dr, "i_{EE'}"'] & A_E \arrow[d, "i_{E'}"] \\
& A_{E'}
\end{tikzcd}$$

(a)

$$\begin{tikzcd}
S \arrow[r, "i_{E'}"] \arrow[dr, "i_{E'E}"'] & A_{E'} \arrow[d, "i_E"] \\
& A_E
\end{tikzcd}$$

(b)

Fig. 1.8. The selective action of E followed by E' (diagram (*a*)) differs from the resulting selective action in the reverse order (diagram (*b*)).

sembled from units supplied by the environment: the latter provides the opportunity for self-assembly, hence for emergence. In sum, the environment of any system is creative – only, it is selective and exclusive rather than permissive.

3.4. *Evolution*

Self-assembly may result in evolution – and this at the molecular, the organismic, the populational, and other levels. Let us then examine the general concept of evolution. To this end let us begin by elucidating the general concept of descent:

DEFINITION 1.16 Let S be a collection of systems of some kind. Then for any x and y in S,

(i) x is an *immediate ancestor* of y (or y *descends immediately* from x) iff x and y belong to the same species and x or a part of x is a precursor in the assembly of y;

(ii) x is a *mediate ancestor* of y (or y *descends mediately* from x) iff there is a z in S such that x is an immediate ancestor of z, and z an immediate ancestor of y;

(iii) x is an *ancestor of* y (or y *descends* from x) iff x is either an immediate or a mediate ancestor of y. Symbol: $x < y$;

(iv) the *ancestry* of x is the collection of ancestors of x:

$$A(x) = \{y \in S \mid y < x\};$$

(v) the *progeny* of x is the collection of things of which x is an ancestor:

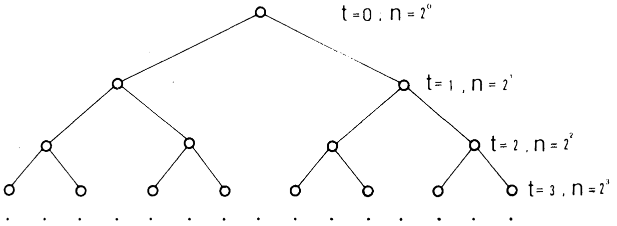
$$P(x) = \{y \in S \mid x < y\};$$

(vi) the *lineage* of x is the union of the ancestry and the progeny of x:

$$L(x) = \{y \in S \mid y < x \quad \text{or} \quad x < y\}.$$

Clearly, the ancestry relation $<$ is a strict partial order. The graph of $<$ is then directed with single edges and no loops.

Example 1 Hydrocarbons descend from C and H, polymers from the respective monomers, ribosomes from RNA molecules and proteins, animals from their parents, etc. *Example 2* The progeny of a bacterium reproducing by binary division looks like this:



$t = 0, \quad n = 2^0$

$t = 1, \quad n = 2^1$

$t = 2, \quad n = 2^2$

$t = 3, \quad n = 2^3$

While in many cases the members of the lineage of a system belong to the same species, sometimes they do not. When they don't, the lineage constitutes an evolutionary line. More precisely, we make

DEFINITION 1.17 Let $L(x)$ be the lineage of a system x of kind K. Then $L(x)$ is an *evolutionary lineage* iff at least one ancestor or one descendant of x belongs to a kind K' different from K.

The notions of descent, lineage and evolutionary lineage carry over to collections of systems:

DEFINITION 1.18 Let S be a family of collections of systems: $S = \{S_i$ is a collection of systems $| \; 1 \leq i \leq n\}$. Then for all S_j and S_k in S,

(i) S_k *descends* from S_j iff every member of S_k descends from some members of S_j: $S_j \prec S_k$;

(ii) the *lineage* of S_j is the family of collections descending from S_j or from which S_j descends:

$$\Lambda(S_j) = \{X \in S \mid X \prec S_j \quad \text{or} \quad S_j \prec X\};$$

(iii) $\Lambda(S_j)$ is an *evolutionary lineage* iff at least one of the ancestors or the descendants of S is included in a species different from that which includes S_j.

Recall now that, according to Postulate 1.4, every system is formed by assembly. Since a system and its precursors belong to different species, it follows that assembly ensues in speciation, i.e. the formation of new species. In other words, we have

THEOREM 1.3 Every concrete system belongs to some evolutionary lineage.

In closing this subsection let us emphasize that the preceding hypotheses are assumed to hold for systems of all kinds – physical, chemical, biological, or social. The peculiarities of systems of these system genera will be studied in succeeding chapters.

4. SYSTEMICITY

4.1. *Integration, Cohesion, Coordination*

Whatever is a system is also a whole but not conversely: an aggregate of independent components is a whole but not an integrated or unitary one. (Compare a living being with its ashes.) Now, systemicity or integration comes in degrees: some systems are more tightly knit than others. The degree of integration depends on the connections or links among a system's components relative to the disintegrating actions of the environment. If the inner couplings are "positive" (or "attractive") and strong, the degree of integration is high; if the links are still positive but weak, the degree of integration is low; and if the links are "negative" (or "repulsive"), there is no systemicity or integration at all. Finally, if some of the links are "positive" while others are "negative", the degree of integration depends on which of them are overriding. For example, a stable atomic nucleus is held together by nuclear forces that overcome the electrical repulsions; and a stable human community is held together by participation in enterprises of common interest, the value of which is greater than that of rivalry or competition – until of course the latter gets the upper hand.

In the case of physical, chemical, and perhaps also biological systems, what measures their degree of integration is their binding energy or, what comes to the same, their dissociation energy. This is the minimal energy required to dissociate the system into its components. It is zero for an aggregate. But such a measure is not universal: it fails to apply to systems where information links play at least as important an integrating role as forces proper – as is the case with social systems. In short, there is no universal measure of the degree of integration or cohesion of a system. However, we can assume the methodological postulate that a measure can be set up for each system genus, or even for each class of models, regardless of the nature of the components of the systems represented by such models.

Example Two things, labeled 1 and 2, form a *linear system* if the components of the state function $\mathbb{F} = \langle F_1, F_2 \rangle$ of the latter satisfy the rate equations

$$\dot{F}_1 = a_{11}F_1 + a_{12}F_2, \qquad \dot{F}_2 = a_{21}F_1 + a_{22}F_2,$$

where the a_{ij} are in general complex numbers. The *degree of integration* or *cohesion* of the system can then be defined as

$$w = \frac{|a_{12}|}{|a_{11} + a_{12}|} + \frac{|a_{21}|}{|a_{21} + a_{22}|}.$$

If the a_{ij} are not constant but time dependent, w itself will depend on time. In general, $w: T \to [0, 1]$.

If $a_{12} = a_{21} = 0$, the components do not form a system; in every other case they do. In particular, if component 1 controls component 2, then $a_{12} = 0$ and $a_{21} \neq 0$, whence $0 < w \leq 1$. And if there is symmetric interaction, $a_{12} = a_{21} \neq 0$. Finally, if all the a_{ij} equal unity, $w = 1$, i.e. the system is maximally cohesive.

Assume then that it is possible to define in each case a measure $w: T \to [0, 1]$ of the degree of integration of systems of any given kind. Then by charting the course of w values we can follow the history of the system from buildup to breakdown through its stable stage if any. In other words, we can introduce

DEFINITION 1.19 Let σ be a system with degree of cohesion or integration $w(t)$ at time t. Then σ is *stable* during the time interval τ iff $w(t)$ is constant for all $t \in \tau$ or at most fluctuates within bounds around a constant value. Otherwise σ is *unstable* and, in particular,

(i) *builds up* (*integrates* or *assembles*) iff its degree of integration increases in time;

(ii) *breaks down* (*disintegrates* or *dismantles*) iff its degree of integration decreases in time.

The degree of integration or cohesion of a system is related to its size or number of components as well as to the nature of the latter. A system with an extremely large number of components may be unstable and ultimately break up into a number of subsystems: there is always some upper limit to the size of a system – a limit to growth. See Figure 1.9. We summarize this empirical generalization into

POSTULATE 1.7 For every kind of system there is an optimal size, i.e. a number of components that maximizes the degree of integration (cohesion) of the system in the given environment. That number is called the *critical size*.

An immediate consequence of this assumption is that, for every kind of system, there is (*a*) a *threshold size*, i.e. a number of components below

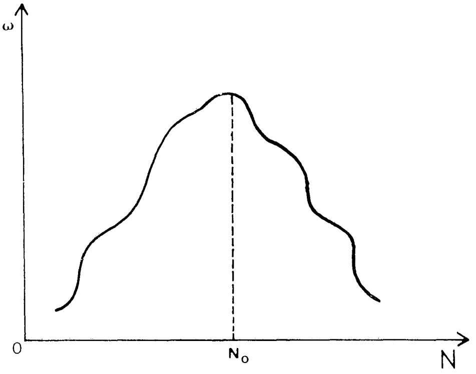


Fig. 1.9. Degree of integration (cohesion) per number of components vs. total number of components. N_0 = optimal or critical size.

which the aggregate does not form a system, and (*b*) a *maximum size*, i.e. a number above which the system breaks down.

Remark 1 Unlike most other axioms of our theory, Postulate 1.7 is empirically testable. For one thing it might well be that for systems of certain kinds there is more than a single critical size. If this were the case we would have to make a small adjustment in Postulate 1.7. *Remark 2* Accretion by gravitational pull would seem to refute our postulate. It does not, for the same process increases the density of matter and radiation energy, and this eventually sets up nuclear reactions that may lead to either explosion or collapse.

The preceding axiom concerns the overall integrity or cohesion or a system without regard for the integration of its subsystems. If a system does have subsystems, not just components, then the cohesion of the subsystems competes with that of the overall system. *Example 1* A molecule with all of its valences saturated is stable, hence it coheres poorly with similar molecules, e.g. its fellow components in a polymer. In general, the greater the complexity of a molecule the lower its overall binding energy. *Example 2* A large human family has weaker interactions with the rest of society, per family member, than a small family: the members of a large family spend a larger amount of energy interacting among themselves than with their social environment. Hence while they fare well in cases of disaster they may not make good citizens.

We compress these observations into

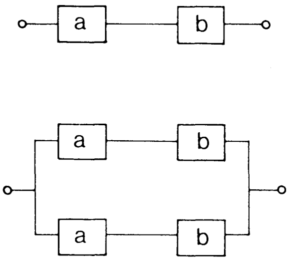
POSTULATE 1.8 The more cohesive each subsystem the less cohesive the total system.

A problem for the systems designer, be he an engineer or an applied social scientist, is to hit on a structure that will maximize overall integrity. He cannot maximize the cohesion of each subsystem because then it becomes self-sufficient instead of serving the whole. And he cannot minimize the partial integrities because then the subsystems would become unstable (unreliable). A compromise solution is to choose medium cohesion subsystems and to have more than one perform a given function or role. Such design enhances the reliability of the system regardless of its nature. See Figure 1.10.

Finally, another concept relevant to that of systemicity is the notion of coordination, which must be distinguished from that of integration. If integration fails the system undergoes *structural breakdown.* On the other hand coordination concerns the relation among either components or functions resulting in functional maintenance. If coordination fails the system undergoes *functional breakdown.* There can be integration without coordination but not conversely. A complex machine out of kilter is integrated but not coordinated. On the other hand organisms are coordinated and *a fortiori* integrated as long as they live. A possible characterization of the concept of coordination is given by

DEFINITION 1.19 If x and y are either in the composition or in the structure of a system, then x and y are said to be *coordinated* iff they contribute jointly to the integrity of the system.

Coordination does not exclude inhibition. Quite on the contrary: when



Reliability $= R = P(a) \cdot P(b)$
Probability of failure $= 1 - R$.

Probability of failure of one branch $= 1 - R$.
Probability of failure of both branches $=$
$(1 - R)^2 \leq 1 - R$
Reliability of system $= 1 - (1 - R)^2 = 2R - R^2$
$\cong 2R$ for small R.

Fig. 1.10. Increasing overall reliability (or degree of integration), by increasing redundancy, i.e. the number of subsystems performing the same function. $P(a)$ is the probability that component a will perform its regular function(s).

coordination is a result of control it includes feedback, which, when negative, is a kind of inhibition. Indeed without such a control excitation might destroy the system. But of course there can be coordination without the intervention of a control system. For example the corpus callosum bridges the two brain hemispheres in vertebrates and thus renders their coordination possible but is not a control system itself. On the other hand the entire central nervous system, which is a controller, coordinates all the subsystems that make up the vertebrate organism.

So much for the concept of wholeness. We now turn to the three main philosophical doctrines concerning wholes.

4.2. *Holism, Atomism, Systemism*

There are three possible doctrines concerning wholes: holism, atomism, and systemism. Holism is the ontological view that stresses the integrity of systems at the expense of their components and the mutual actions among them. It is characterized by the following theses.

H1 The whole precedes its parts. At first blush this thesis looks true. Thus before going for a hair cut we must have grown some hair. But of course the hair grew gradually, not all of a sudden: it became a whole in the course of a process of multiplication of parts (cells). Before making a general pronouncement concerning which precedes which we should examine the actual process in question. A system precedes its components only during a breakdown process; it succeeds them during the synthesis or formation process. In any case the existence of a system may not be obvious: it may call for an explanation in terms of both the mutual actions of parts and environment. No such explanation will be sought if the whole is taken for granted and regarded as the ultimate ground for the existence of its parts.

H2 The whole acts on its parts. For example, it will be said that the needs of the organism (or the society) as a whole dictate the functioning of its parts. But of course there would be no whole were it not for the coordination of its parts. There is no action of the whole on ist parts; rather, there are actions of some components upon others. Thus the vibration modes of any single particle in an elastic body are influenced by the motion of the other particles; likewise the behavior of any person is partially determined by that of his fellow society members. In all these cases we have not the whole acting on its parts but some or even all of the remaining components of the system acting on the given component, or the behavior of the latter being partly determined by the place it occupies in the system, in particular by its function or role.

H3 The whole is more than the sum of its parts. As it stands the thesis is hardly intelligible. It becomes intelligible if by 'sum' one means the juxtaposition (physical sum or association $+$) that we met in Sec. 1.6, and if by 'more' one means that the whole, provided it is a system, has emergent properties that its components lack (cf. Sec. 3.2). Reformulated in this way, i.e. in a nonholistic fashion, H3 acquires a definite sense – but it proves to be only partially true. Indeed, although every system is a whole, not every totality is a system; thus the mere aggregation of things need not result in an integrated whole or system (cf. Sec. 4.1.). What make a whole into a system are precisely the actions exerted by some of its parts upon others. But the holist does not care for the disclosure of such couplings, i.e. the system structure: he despises analysis.

H4 Wholes emerge under the action of agents that transcend both the actions among the components and the environmental influences. For example, morphogenesis is guided by an entelechy, or *élan vital*, or morphogenetic field external to the components. In sum, the formation of totalities transcends their components and is traceable to inscrutable entities. Thus far the holistic account of the formation of wholes. Needless to say, science has no use for such secret hence incontrollable principles of organization. Instead, science acts on a principle of immanence, not transcendence, namely this: Only the components, the way they come together, and the environment determine what kind of thing a totality shall be. (Hence our representing a system by the ordered triple: composition-environment-structure.)

H5 Totalities cannot be explained by analysis: they are irrational. This thesis is trivially true if by 'analysis' is meant only decomposition into parts, since then only the composition of a system but not its structure is revealed. If the latter is left out then of course it becomes impossible to account for the systemic or gestalt properties of a totality. But the physicist does not claim that water is just an aggregate of H_2O molecules, and the sociologist does not assert that society is just a collection of persons. In either case the links among the components (hydrogen bonds, work relations, or what not) must be disclosed or hypothesized to understand the formation, cohesion, and eventual breakdown of a totality. Such an analysis is the conceptual basis for any effective synthesis or upward building, as well as for the effective analysis or disintegration of a system.

H6 The whole is better than any of its parts. This value judgment has been brandished as an ideological hatchet aimed at suppressing the rights of individuals and groups in the name of the good of the whole or higher system, in particular the economico-political establishment, regardless of whether the latter does in fact work for the common good. We shall waste no time on this fig leaf.

Holism, in a nutshell, is antianalytic and therefore antiscientific. In effect, it has been responsible for the backwardness of the nonphysical sciences. And it has contributed precious little to serious systemics precisely because (*a*) it has not engaged in a study of the links that hold any system together, and (*b*) rather than constructing conceptual systems (theories) to account for concrete systems, it has spent itself in attacking the analytical or atomistic approach and praising totality as such. Whatever truth there is in holism – namely that there are totalities, that they have properties of their own, and that they should be treated as wholes – is contained in systemism, or the philosophy underpinning systemics or the general theory of systems (cf. Bunge, 1977d).

In opposing holism we do not espouse its opposite, namely atomism – the thesis that the whole is somehow contained in its parts, so that the study of the latter should suffice for understanding the former. Certain wholes, to wit systems, do have collective or systemic properties not possessed by their components, and this is why they must be studied as systems. Consider the celebrated though ill understood example of a so-called contingent identity, namely Water = H_2O. This is no identity at all for, whereas the LHS is short for 'water body' (e.g. a lake), the RHS describes a property of its molecular components. (No identity is possible between a thing and a property.) What is true, of course, is that the molecular composition of water is a set of H_2O molecules, but this is no identity statement. (In other words, the correct statement is this: For any water body w, $\mathcal{C}(w) \subset$ The set of H_2O molecules.) Moreover, specifying the composition of a system does not suffice to characterize it as a system: we must add a description of the system structure. And it so happens that water, as a system composed of myriads of H_2O molecules, has properties that none of its components has – e.g. transparency, a high dielectric power (hence a high dissolving power), freezing at 0°C, and so on and so forth. Some such properties must be included in any realistic model of water.

The ontological differences between a body of water and a H_2O molecule are such that, to account for the behavior of the former, we need not only all the knowledge we have about the individual H_2O molecule but also a host of hypotheses and data concerning the structure of water (i.e. the relative configuration of H_2O molecules in the lattice) as well as hypotheses and data about the dynamics of water bodies – hypotheses and data which vary of course according to whether water is in the gaseous, liquid, or solid phase. In sum, to describe, explain or predict the properties of water we use both microlaws and macrolaws.

Atomism, an ontological doctrine, is usually, though not necessarily, allied with *reductionism*, the epistemological doctrine according to which the study of a system is reducible to the study of its components. (The converse is false: one can be an epistemological reductionist, yet acknowledge wholes, emergence, and levels.) The reductionist will claim of course that we may use macrolaws and, in general, laws of systems, as a convenience, though in principle we should be able to get along with microlaws (or laws of components) alone, since the former are reducible to (deducible from) the latter. This thesis contains a grain of truth but is not the whole truth. No theory T_2 of water as a body follows solely from a microphysical theory T_1 of the H_2O molecule – not even by our adjoining what some philosophers call the bridge laws relating macrophysical concepts (e.g. pressure) to microphysical ones (e.g. molecular impact). Much more than this must added to the primary or reducing theory T_1 in order to obtain the secondary or reduced theory, namely hypotheses concerning the interactions among the system components.

The extreme case of reduction is that of straightforward deduction from a given set of premises, or *strong reduction*. Examples: reduction of particle mechanics to continuum mechanics, and of ray optics to wave optics. (The converse reductions are impossible.) These are rather exceptional cases. In general we must resort to a more complex strategy, namely *weak reduction*, or deduction from a primary theory in conjuction with a set of conjectures and data congenial with but alien to the former. The structure of this inference is:

$$T_1 \cup \text{Subsidiary hypotheses and data concerning interactions among components} \vdash T_2.$$

The subsidiary hypotheses constitute a model of the composition and structure of the system. Since this model, though couched in the language of T_1, is not included in T_1, we are in the presence not of straight (or strong) but of partial (or weak) reduction. (Cf. Bunge, 1977f.)

Note that we are not stating that the properties of water, or of any other macrosystem, are mysterious. On the contrary, they can be explained at least in outline. For example, the exceptionally high boiling point and evaporation heat of water are explainable in terms of the hydrogen bonds linking together all the H_2O molecules in a water body, bonds which are in turn explained by the composition and the structure of the H_2O molecule. But the point is that the intermolecular hydrogen bonds do not occur in the study of the individual H_2O molecule. In other words, although water is composed of H_2O molecules it does not reduce to H_2O – *pace* the efforts of enlightened philosophers to crush the holistic monster (e.g. Kemeny and Oppenheim, 1956; Oppenheim and Putnam, 1958, Putnam, 1969).

In short, atomism is almost as false as holism, the difference lying in that, whereas the former stimulates research, the latter blocks it. Each of these views has a grain of truth that the systemic view preserves and expands.

5. CONCLUDING REMARKS

The idea of a system, as distinct from a loose aggregate, is very old. Yet it has only recently been elucidated and exploited systematically. The mere suggestion that the thing we happen to be looking at, manipulating, or investigating might be a system rather than an unstructured object or a mere heap, will guide our study and handling of it. Indeed, if we suspect that a certain thing is a system then we shall endeavor to identify its composition, its environment, and its structure.

The order in which the three coordinates of the system concept appear is natural rather than accidental. Indeed, listing a system's components must precede any question regarding its environment and its structure; and identification of the environment comes before display of the structure, because the latter is the collection of relations among the components and among these and environmental items. True, when coming across certain systems, such as a plant, a watch, or a galaxy, we often begin our search with the totality and its environment, ending up by disclosing its composition and its structure. But when investigating a forest, a social system, and *a fortiori* a social supersystem such as a nation, we first meet its components (or atomic parts) in their milieu and try to figure out the structure of the whole by studying the behavior of the individual components. In either case, i.e. whatever our mode of perception may be, the conceptual analysis of a system must proceed in the way indicated – identification of composition, environment, and structure – if only for mathematical reasons.

Indeed it makes no sense to hypothesize any relations without knowing what the relata (system components and environmental units) may be. Therefore the holistic claim that atomistic analysis or the part method is incapable of grasping totalities, is groundless. On the contrary, holism is incapable of accounting for any totalities precisely because it refuses to disclose the components that are held together in the system: no components, no bonds among them. This is not to condone the antithesis of holism, namely atomism, and its usual epistemological partner, reductionism, according to which wholes – in particular, systems – are artifacts, emergents are identical with resultants, and levels are just convenient methodological categories.

The world view that emerges from this chapter is systemic: it holds that the universe is a system composed of subsystems. More precisely, the universe is the supersystem of all other systems. The world is thus neither a solid block nor a heap of disconnected items. It is held together by a number of links, from intermolecular bonds to gravitation to information. The world is material but not just a lump of physical entities: it is composed of systems of a number of qualitatively different kinds. Although all systems are physical not all of them are just physical. The universe is enormously variegated: its components can be grouped into a number of levels, such as the physical, the chemical, the biological, and the social levels. Moreover the world is restless and all of its changes are patterned (lawful). The world is, in sum, a coherent or integrated system of systems, and one that is varied, changeable, and regular.

CHAPTER 2

CHEMISM

All concrete systems are physical entities, in that they possess energy – a measure of changeability (Vol. 3, Postulate 5.3). However, some concrete systems are more than just physical entities, in the sense that the categories of physics are insufficient to account for them. For example, some are social systems. And although some of the concepts of physics – in particular the notion of energy – are necessary to account for any society, most are unnecessary. On the other hand social science requires new concepts, such as those of artifact, social group, justice, and culture. The peculiar concepts of social science are supposed to represent emergent properties of a society, i.e. properties that are not possessed by its components. Likewise the study of the components of any society, i.e. social animals, calls for a number of distinctly biological categories, such as reproduction, sex, and death. And if the animals happen to have a highly evolved nervous system, we shall need still other categories, such as the concepts of altruism and planning, which are usually regarded as psychological. Finally, organisms are composed of chemical systems, which are physical entities of a very special kind, namely composed of atoms and molecules engaged in chemical reactions.

Thus the totality of concrete entities may be grouped into five genera: the set of physical things (e.g. atoms and fields), the set of chemical systems (e.g. protein-synthesizing units), the set of biosystems (organisms and their organs), the set of sociosystems (societies and their subsystems and supersystems), and the set of artificial things. (The set of psychosystems, or animals endowed with mental abilities, is assumed to be a subset of that of biosystems.) The first three genera precede the other two in the sense that they supply the components of the higher order things or take part in some other way in their assembly. Since the lower a genus the more populous it is, we may depict the structure of reality as a pyramid: see Figure 2.1. This suggests the plan of this volume. (More on this structure in Ch. 6.)

In this chapter we shall hardly touch on physical systems: first, because they are the best understood of all; second, because we have actually dealt with them in Vol. 3. We shall focus instead on certain key chemical categories of crucial importance in the study of life, such as those of che-

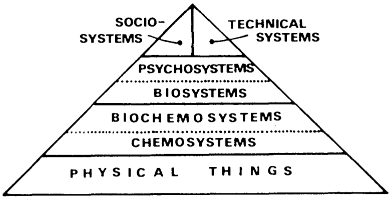


Fig. 2.1. The five main genera of things constituting the furniture of the world.

mical reaction and chemosystem, protein and nucleic acid. These categories are certainly more specific (less universal) than those of thing, property, system, and others we have encountered so far. On the other hand our chemical categories are far more universal than some of those that have dominated ontology for centuries, such as those of mind and purpose.

1. CHEMICAL SYSTEM

1.1. *Atom*

Since we shall make ample use of the concept of a chemical system, we must characterize it, the more so since no definition of it in systems theoretic terms seems to be available. And, since a chemical system is one whose components are reacting chemicals, we must start by characterizing the notions of a chemical and of a reaction.

The components of a chemical system are chemical atoms or molecules. A detailed characterization of things of these kinds calls for a number of specific scientific concepts. Luckily atoms have an essential or root property from which many another depends, namely atomic number. And molecules have an essential property of their own, namely atomic composition. These two properties will suffice for our limited purposes.

Our starting point is then the notion of chemical atom of a kind, or elementary chemical species, or element for short. There are 100 odd elements, among them hydrogen (H), carbon (C), nitrogen (N), and oxygen (O), the atomic numbers of which are 1, 6, 7, and 8 respectively. Every member of such a natural kind or species is a material system formed by at least two things of different genera, namely a system of nucleons (protons or neutrons), and an aggregate of electrons. In turn the nucleons are probably composed of subunits (quarks?), and the atoms can join forming molecules. See Figure 2.2. (This diagram suggests that molecules and their

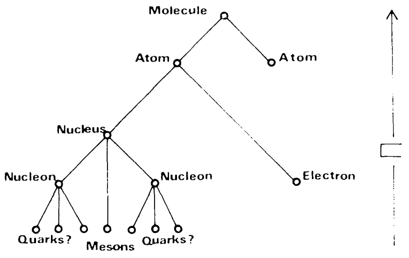


Fig. 2.2. A molecule is composed of two or more atoms. Each of these is in turn composed of a nucleus and one or more electrons. The number of protons in the nucleus equals the atomic number of the atom. Nucleons are possibly held together in the nucleus by meson fields. In turn every nucleon (proton or neutron) is presumably composed of smaller units (quarks?). Altogether six different kinds of thing are involved in a molecule.

components constitute a *sup* semilattice the order of which is that of the part-whole relation $\sqsubset$. Thus if a_1 and a_2 are two atoms forming a molecule m, then $sup\ \{a_1, a_2\} = m$. However, we shall not pursue this point here.)

In short, we steal from physics and chemistry

POSTULATE 2.1 Let A be a nonempty subset of the collection of all concrete things (or entities), and $\mathbb{N}_0$ the set formed by the first N positive integers. There exists a many to one relation (function) $Z : A \to \mathbb{N}_0$ of A onto $\mathbb{N}_0$(i.e. a surjection) that assigns each thing in A a natural number in $\mathbb{N}_0$, representing a property of the thing.

A is of course the set of atoms and Z the atomic number. The atoms with the same Z value are Z-equivalent:

If x and y are in A, then $x \sim_Z y =_{df} Z(x) = Z(y)$. The quotient of A by $\sim_Z$, or $A/\sim_Z$, is the family of classes of atoms with the same Z value. (We do not need to make the finer distinction among the isotopes of an element.) In other words, we can make

DEFINITION 2.1 Let A be the domain of the function $Z: A \to \mathbb{N}_0$. Then

(i) if x is in A, x is an *atom*;

(ii) the property represented by the function Z is called the *atomic number*, and the value $Z(a)$ of Z at $a \in A$, the *atomic number* of a;

(iii) the nth set

$$^{n}A = \{x \in A \mid Z(x) = n\}, \quad \text{for } 1 \leq n \leq N,$$

is the nth *atomic kind* or *chemical element*;

(iv) the partition

$$\mathscr{A} = A/\sim_Z = \{{}^nA \subset A \mid 1 \leq n \leq N\}$$

is the *family of atomic kinds* or *chemical elements*.

Since by definition $\mathscr{A}$ is the partition of A induced by $\sim_Z$,

$$A = \bigcup_{n=1}^{N} {}^nA, \quad \text{and} \quad {}^mA \cap {}^nA = \varnothing \quad \text{for } m \neq n.$$

In other words, the totality of atoms is the disjoint union of all the atomic kinds. The latter are those displayed in the periodic table. It is not known whether the number of atoms of each atomic kind is finite.

So much for the (atomic) components of a molecule.

1.2. *Molecule*

A molecule is a system composed of atoms – or, if preferred, of nuclei and electrons. Molecules range from hydrogen or H_2 (molecular weight $= 2$) to deoxyribonucleic acids or DNA (molecular weight up to 1 billion). There are about 4 million known species of molecules, 96 percent of which contain carbon. The combinations of atoms to form molecules are brought about by either ionic bonds or covalent bonds. In turn molecules stick together by virtue of hydrogen bonds or van der Waals forces (e.g., hydrophobic forces). All of these bonds are ultimately electric forces and they are studied by quantum chemistry, an application of quantum mechanics.

In ontology we take such bonds for granted and let chemists inquire into their nature. (Remember the fate of Hegel's alleged refutation of Berzelius' brilliant hypothesis that chemical bonds are electrical: cf. Hegel (1830), Sec. 330). The metaphysician accepts gratefully what the chemist is willing to teach him about the spontaneous formation (self-assembly) of molecules out of atoms, or of other molecules, as well as about the breakdown or dissociation of such systems as a result of their interaction with their environment (in particular thermal motion). The philosopher's task is not to compete with the chemist but to try and discover the structure and peculiarities of chemical processes *vis à vis* other kinds of process. If in the course of his work he comes up with results helping to clarify the foundations of chemistry, so much the better.

The statement that all molecules are composed of atoms can be exactified in a number of ways. We shall proceed as follows. Consider for example the first atomic species or

$$^1A = \mathrm{H} = \text{The set of all hydrogen atoms} = \{x \in \Theta \mid Z(x) = 1\}.$$

Since we disregard the individual differences among the hydrogen atoms, to concentrate on their shared atomic number, we may regard the species H as formed by an unspecified number of copies of a single individual. Assume now that each hydrogen atom can combine with any number of other H atoms forming hydrogen molecules – possible ones like the diatomic ones, and impossible ones like the icosiatomic ones. Since we are disregarding idiosyncrasies and concentrate on atomic composition, we can speak of the m-fold concatenation of an unspecified hydrogen atom x with itself, or $x \cdot x \cdot \ldots \cdot x = x^m$, as representing the combination of m hydrogen atoms. In this form we fashion

$$^1A^* = \mathrm{H}^* = \text{The set of finite concatenations of hydrogen atoms.}$$

Imagine now that the same construction is carried out for every atomic species, so that we have

$$^nA^* = \text{The set of finite concatenations of atoms of kind } {}^nA.$$

This set is the union of the following sets:

$$\begin{aligned} {}^nA_1 &= \{z \in \Theta \mid z \in {}^nA\} \\ {}^nA_2 &= \{z \in \Theta \mid z = x^2 \;\&\; x \in {}^nA\}, \\ &\ldots\ldots\ldots\ldots\ldots\ldots \\ {}^nA_m &= \{z \in \Theta \mid z = x^m \;\&\; x \in {}^nA\}. \end{aligned}$$

In other words, the set of all conceivable ^nA compounds is

$$^nA^* = \bigcup_m {}^nA_m = \{z \in \Theta \mid z = x^m \;\&\; x \in {}^nA \;\&\; 1 \leq m < \infty\}.$$

But of course not all conceivable compounds are really possible, i.e. lawful. (For the definition of real possibility in terms of lawfulness, see Vol. 3, Ch.4, Sec. 2,3.) The really possible compounds from ^nA are a proper subset of $^nA^*$, i.e.

$$^nA^{**} = \{z \in \Theta \mid z = x^m \;\&\; x \in {}^nA \;\&\; 1 \leq m < \infty \;\&\; z \text{ is lawful}\}.$$

(Do not think just of the stable compounds satisfying the valence requirements but also of all the transient molecules and free radicals.) These are all the homogeneous compounds, i.e. the molecules composed of atoms of a single atomic species. A heterogenous compound, such as the member of the chemical species CH_4, or methane, is a member of the set

$${}^1A_4{}^6A_1 = \{z \in \Theta \mid z = x^4y \;\&\; x \in {}^1A \;\&\; y \in {}^6A\}.$$

The totality of heterogeneous compounds is

$$A^* = \{z \in \Theta \mid z = x_1^{m_1} x_2^{m_2} \ldots x_N^{m_N} \;\&\; x_i \in {}^iA \;\&\; 1 \leq i \leq N \;\&\; 0 \leq m_i < \infty\}.$$

This is the set of all conceivable molecules, in turn a superset of the set of all really possible molecules.

The structure of A^* is obviously that of a semigroup under the binary operation $\cdot$ of combination or concatenation. More exactly, the structure $\langle A^*, \cdot\rangle$ is the *free semigroup generated by the set A* of atoms, and this because every element of A^* is either a member of A or a string of members of A. Now, as noted before, A^* contains not only the really possible molecules but also the unlawful ones. If we shrink A^* to the subset A^{**} of really possible molecules, we face the difficulty that A^{**} is not closed under the operation $\cdot$ of combination or concatenation (i.e. it does not contain all the concatenates). Equivalently: $\cdot$ is a *partial* operation on A^{**}, i.e. one not defined for every pair of members of A^{**}. We shall therefore say that the system $\mathscr{A}^{**} = \langle A^{**}, \cdot\rangle$ is a *partial* semigroup *embedded* in the free semigroup $\mathscr{A}^* = \langle A^*, \cdot\rangle$ generated by A. In other words, we make

POSTULATE 2.2 Let $\mathscr{A}^* = \langle A^*, \cdot\rangle$ be the free semigroup generated by the set A of all atoms of all kinds (Postulate 2.1), and call

$$A^{**} = \{z \in \Theta \mid z = x_1^{m_1} x_2^{m_2} \ldots x_N^{m_N} \;\&\; x_i \in {}^iA \;\&\; 1 \leq i \leq N \;\&\; 0 \leq m_i < \infty \;\&\; z \text{ is lawful}\}.$$

Then

(i) the partial semigroup $\mathscr{A}^{**} = \langle A^{**}, \cdot \rangle$ is embedded in $\mathscr{A}^*$;

(ii) for any two things x and y in A^{**}, if xy is in A^{**}, then xy represents the combination of x with y (or the xy compound).

Note that we are not postulating that $\mathscr{A}^{**}$ is a commutative semigroup. Hence the isomers xyz and yxz will be distinct molecules.

The previous axiom supports the following conventions that will come in handy in the sequel.

DEFINITION 2.2 Let $\mathscr{A}^{**}$ be the set defined in Postulate 2.2 Then

(i) the set of *compounds* is the complement of A^{**} to A, i.e.

$$M = A^{**} - A = A^{**} \cap \bar{A};$$

(ii) any subset of M, of the form

$$M_j = \{z \in \Theta \mid z = x_1^{m_1} x_2^{m_2} \dots x_N^{m_N} \ \& \ x_i \in {}^iA \ \& \ 1 \leq i \leq N\},$$

with fixed values of the m_i for $0 \leq m_i < \infty$, is a *molecular species*;

(iii) every member of a molecular species is a *molecule.*

(Note the strategy of first defining a set and then naming its members.)

DEFINITION 2.3 The *composition* of a molecule is the set of its constituents. I.e. let

$$\mathscr{C}^{**} : A^{**} \longrightarrow 2^{A^{**}}$$

be the function from the set of atoms and molecules to its power set, such that, if $z = x_1^{m_1} x_2^{m_2} \dots x_N^{m_N} \in A^{**}$, where $x_i \in {}^iA$ and $0 \leq i < \infty$,

$$\mathscr{C}^{**}(z) = \{x_1^{m_1}, x_2^{m_2}, \dots, x_N^{m_N}\}.$$

Then $\mathscr{C}^{**}(z)$ is the *composition* of the molecule z.

Example The composition of glucose, or $C_6H_{12}O_6$, is given by:

If $z \in C_6H_{12}O_6$, then $\mathscr{C}^{**}(z) = \{6$ carbons, 12 hydrogens, 6 oxygens$\}$.

The next definition requires a generalization, to an arbitrary set T of things, of the notion of physical addition (or association) occurring in the formula '$a + b = c$.' For an arbitrary set T of things, the association or addition of its members is the thing $[T] = \sup T =$ least upper bound of T. (Clearly, for any x in T, $x \sqsubset [T]$.)

DEFINITION 2.4 Let x be a thing. Then

(i) x is a *chemical* iff x is either a member of A^{**} (i.e. an atom or a molecule) or the aggregation of a finite set of atoms or molecules – i.e. if there is a $T \subset A^{**}$ such that $x = \sup T = [T]$;

(ii) x is a *compound* iff x is either a molecule or the aggregation of a finite set of molecules, i.e. $x \in M = A^{**} - A$ or $x = [T]$, with $T \subset M$;

(iii) x is a *polymer* iff x is either a string of molecules of a single molecular species, or the aggregation of a finite set of strings of different molecular species.

Polymers are of particular interest to us because all biomolecules are polymers. In particular, the DNA molecules are extremely long chain polymers.

Finally we make

DEFINITION 2.5 Let A^{**} be the totality of compounds and $l: A^{**} \to \mathbb{N}$ a function assigning each atom or molecule a natural number. Then the *length* of a member z of A^{**} equals the number of atoms in z. More precisely,

$$\text{If } z = x_1^{m_1} \cdot x_2^{m_2} \ldots x_N^{m_N} \in A^{**}, \quad \text{then } l(z) = m_1 + m_2 + \cdots + m_N.$$

This concept of length applies to molecules of all shapes, not just the linear ones. Thus the length of the planar molecule CH_4 is, according to our definition, equal to 5.

So much for structural matters. (For somewhat different treatments see Rosen (1959) and Chiaraviglio (1965).) Let us now turn to chemical processes.

1.3. *Chemical System*

We shall characterize a chemical system as a system whose components are chemicals varying in number (or in concentration) for being engaged in reactions. To render this idea more precise we shall avail ourselves of the state representation of systems explained in Ch. 1, Sec. 2.2. Consider a concrete system represented by a time-dependent state function $\mathbb{F} = \langle F_1, F_2, \ldots, F_n \rangle$, each component (coordinate) of which represents a property of the system. The state function spans the total lawful state space of the system. Now single out those components of $\mathbb{F}$ that represent the numbers (or else the concentrations) of the chemicals in the system. The resulting state function, and the accompanying state subspace, deserve names of their own:

DEFINITION 2.6 Let T be the time span (relative to a given reference frame) and let

$$\mathbb{F} = \langle F_1, F_2, \ldots, F_n \rangle : T \longrightarrow \mathbb{R}^n$$

be a state function for an arbitrary system σ of a certain kind, satisfying the laws in a set $\mathbb{L}$. Further, let the first $m < n$ coordinates of $\mathbb{F}$ (at a given time) equal the number of components of each kind of chemical in the system (at that time). Then

(i) the function

$\mathbb{F}_c = \langle F_1, F_2, \ldots, F_m \rangle$, with $F_i(t)$ = Number of atoms or molecules of kind $K_i \subset A^{**}$ in σ at time $t \in T$, is called the *composition state function* of σ;

(ii) the value of the composition state function at time $t \in T$, i.e. $\mathbb{F}_c(t)$, is called the *instantaneous composition* of σ;

(iii) the set of all possible instantaneous compositions of σ, i.e.

$$S_c(\sigma) = \{\mathbb{F}_c(t) \in \mathbb{R}^n \mid t \in T \ \& \ \mathbb{F}_c \text{ satisfies } \mathbb{L}\},$$

is called the *composition state space* of σ.

Example The composition state function of a system where water is formed out of hydrogen and oxygen, according to the reaction $2H_2 + O_2 \rightarrow 2H_2O$, has

$$\text{the } \textit{initial value } \mathbb{F}_c(t_i) = \langle 2, 1, 0 \rangle$$

and (on completion of the reaction)

$$\text{the } \textit{final value } \mathbb{F}_c(t_f) = \langle 0, 0, 2 \rangle,$$

where the integers are the numbers of molecules of the chemicals in the system. If we wish to represent the entire process from start to end, rather than only the initial and the final states, we can use the partial concentrations, to obtain

$$\mathbb{F}_c = \langle C_{H_2}, C_{O_2}, C_{H_2O} \rangle,$$

where each coordinate of the triple is a function from T to $\mathbb{R}$. A condition for being able to use these functions is that the system be rather bulky, as is tacitly assumed in chemical kinetics.

Finally we can make

DEFINITION 2.7 Let σ be a concrete system with an n component state function $\mathbb{F}$ (in a given representation). Then σ is a *chemical system*, or *chemical reactor*, during the period $\tau \subseteq T$, iff

(i) the first m coordinates of $\mathbb{F}$, where $0 < m < n$, form the composition state function $\mathbb{F}_c$ of σ (i.e., σ is composed, among other things, of chemicals: $\mathscr{C}(\sigma) \cap A^{**} \neq \varnothing$);

(ii) the instantaneous value of the ith component F_i of $\mathbb{F}_c$, for $1 \leq i \leq m$, equals the number of atoms or molecules (or else the concentration) of ith chemical natural kind or species included in A^{**} at the time concerned;

(iii) the chemical composition of σ (i.e. the value of $\mathbb{F}_c$) changes in the course of τ, and this change is not just a transport from or to environment of σ.

According to this definition there can be talk of a chemical system only where and while chemical reactions are in progress. Thus a dead battery is no longer a chemical system. Before the reactions start, and after they have reached completion, the system is a physical one and only potentially a chemical system. (A chemical system is then an essentially dynamical system that can be at most in a state of dynamical equilibrium, or steady state, which is the one at which every reaction proceeds at the same pace as its converse, so that there is no net variation of the composition of the system.) This suggests

DEFINITION 2.8 A concrete system σ is a *physical system* during the period τ iff the composition state function of σ is constant (e.g. zero) throughout τ or, if it changes, it does so by transport from or to the environment of σ.

In sum a chemical system is a system whose chemical composition changes in the course of time, however slowly, and does so not just by importation or exportation of chemicals, but by virtue of processes involving its components. These processes are of course the chemical reactions. Let us look into these.

1.4. *Chemical Reaction*

Every really possible (lawful) process in a chemical system is representable as a curve in the total state space of the system. Each m-tuple of initial concentrations of the chemicals in the system corresponds to a distinct trajectory of the tip of the state vector $\mathbb{F}$. ($\mathbb{F}$ is not really a vector, for its product with an arbitrary real number may lie outside the bounds set by the laws.) And the bundle of all such possible trajectories, for all variations of initial conditions and all systems of the same kind, represents the totality $\mathbb{L}$ of laws "at work" in chemical systems. Since a system is a chemical system only as long as there are chemical transformations going on in it, the set $\mathbb{L}$ includes every one of the laws concerning such transformations, in particular the reaction formulas and the corresponding kinetic equations. But, of course, $\mathbb{L}$ includes also every other law satisfied (possessed) by the chemical system of the given kind. See Figure 2.3.

We spell out the foregoing in

DEFINITION 2.9 Let $\mathbb{F}$ be a state function for a concrete system σ of kind K, and let $S_{\mathbb{L}}(K)$ be the corresponding lawful state space. Then

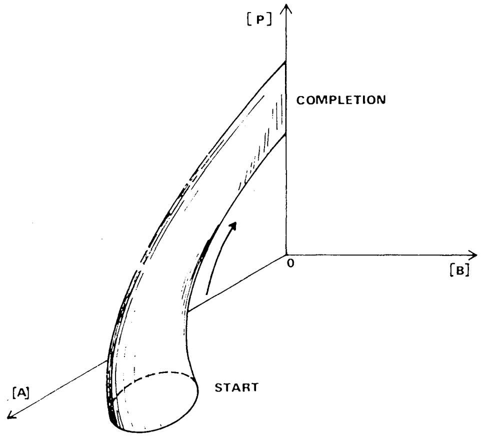


Fig. 2.3. Composition state space of a chemical system in which only reactions of the form $A + B \rightharpoonup P$ occur. $[X]$ designates the concentration of X. Each line within the conoid, from beginning to end, represents one possible complete reaction compatible with the reaction schema as well as with the formula for the rate of formation of P from A and B.

(i) the full trajectory of the state point $\mathbb{F}(t)$ in $S_{\mathbb{L}}(K)$ is called the total *history* of σ during $\tau \subseteq T$:

$$h(\sigma, \tau) = \{\langle t, \mathbb{F}(t)\rangle \mid t \in \tau\};$$

(ii) the set of *laws* of concrete systems of kind K is the family of all the histories of all the systems in K over all periods:

$$\mathbb{L} = \{h(\sigma, \tau) \mid \sigma \in K \,\&\, \tau \in 2^T\},$$

where 2^T is the family of all time periods included in T.

We are interested in certain segments of the history of an arbitrary chemical system called 'chemical reactions'. Equivalently: we shall pay particular attention to the chemical law statements (formulas) that focus on chemical transformations, especially those that disregard all the concomitant physical processes. Consider for example the substitution schema

$$r = \ulcorner A + B \rightharpoonup C + D\urcorner,$$

usually interpreted as "One unit of species A reacts with another of species B to produce one unit of species C and another of species D". (The interpretation of the more general reaction schema $mA + nB \rightharpoonup pC + qD$ is obvious.) This statement says nothing about the reaction mechanism or about the environmental conditions for it. (The latter are taken care of in the antecedent of the conditional statement of which r is the consequent. The entire conditional is then of the form: "If the pressure, temperature, etc., are such and such, then r".) Nor does the statement inform us about the pace of the chemical process: it just tells us what there is in the beginning and at the end of the transformation. I.e. it speaks only of net events. (Moreover r is ambigous; it uses class symbols, such as 'A' and 'B', to refer to members of such classes.) In any case, this is how the above reaction schema can be translated into the state space language: the system is initially in the composition state

$$\mathbb{F}_c(t_i) = \langle C_A, C_B, 0, 0\rangle$$

and finally in the composition state

$$\mathbb{F}_c(t_f) = \langle 0, 0, C_C, C_D\rangle$$

so that the size of the net change is

$$r = \mathbb{F}_c(t_i) - \mathbb{F}_c(t_f) = \langle C_A, C_B, -C_C, -C_D\rangle.$$

Example The overall photosynthesis reaction (which has a great many intermediate steps) is

$$6CO_2 + 6H_2O \rightharpoonup C_6H_{12}O_6 + 6O_2.$$

Setting

$$\mathbb{F}_c = \langle C_{CO_2}, C_{H_2O}, C_{C_6H_{12}O_6}, C_{O_2}\rangle,$$

we get

$$r = \langle 6, 6, 0, 0\rangle - \langle 0, 0, 1, 6\rangle = \langle 6, 6, -1, -6\rangle.$$

And the converse reaction, which (barring differences in intermediate steps) is that of respiration, is of course

$$-r = \langle -6, -6, 1, 6\rangle,$$

a systems theoretic rewrite of

$$C_6H_{12}O_6 + 6O_2 \rightharpoonup 6CO_2 + 6H_2O.$$

We compress the foregoing into the following semantic assumption. Let $\mathbb{F}_c$ be a composition state function for a chemical system σ between the initial time t_i and the final time t_f (relative to a certain reference frame). Then

(i) the *overall* or *net reaction* in σ between t_i and t_f is representable by the difference between the initial and final values of $\mathbb{F}_c$:

$$r = \mathbb{F}_c(t_i) - \mathbb{F}_c(t_f);$$

(ii) while the positive components of r represent the *reactants*, the negative ones represent the *reaction products*;

(iii) the *inverse* (or *converse*) of a reaction r is the reaction $-r$;

(iv) the *null reaction* is $\mathbb{O} = \langle 0, 0, \ldots, 0 \rangle$.

As we know from elementary chemistry, reactions can add to form further reactions: see Figure 2.4. For example, the enzymatic reaction sequence (or pathway) represented by the system of reaction schemata

$$\begin{aligned} &A + C \rightharpoonup AC \\ &AC + B \rightharpoonup ACB \\ &ACB \rightharpoonup AB + C \end{aligned}$$

adds up to the overall reaction

$$A + B \rightharpoonup AB.$$

In our state space representation we form the composition state function

$$\mathbb{F}_c = \langle C_A, C_B, C_C, C_{AB}, C_{AC}, C_{ACB} \rangle,$$

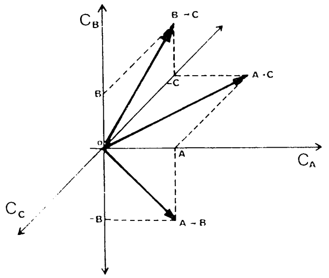


Fig. 2.4. The reactions A ⇀ B and B ⇀ C add up to the overall reaction A ⇀ C. Each reaction is represented by an arrow in the cartesian three-space of the concentrations (or the numbers of atoms or molecules).

so that the three successive reactions are

$$\begin{aligned} r_1 &= \langle 1, 0, 1, 0, 0, 0\rangle - \langle 0, 0, 0, 0, 1, 0\rangle = \langle 1, 0, 1, 0, -1, 0\rangle \\ r_2 &= \langle 0, 1, 0, 0, 1, 0\rangle - \langle 0, 0, 0, 0, 0, 1\rangle = \langle 0, 1, 0, 0, 1, -1\rangle \\ r_3 &= \langle 0, 0, 0, 0, 0, 1\rangle - \langle 0, 0, 1, 1, 0, 0\rangle = \langle 0, 0, -1, -1, 0, 1\rangle. \end{aligned}$$

The overall reaction is formed by adding the coordinates term by term:

$$r = r_1 + r_2 + r_3 = \langle 1, 1, 0, -1, 0, 0\rangle,$$

which, in effect, represents ⌜A + B ⇀ AB⌝. That is, the catalyzer C and the intermediate products cancel out, so that only the consumed reactants and the newly formed chemicals occur in the overall reaction. If the net result had been a return to the initial state of the system, we would have had the null overall reaction.

We summarize the preceding into

DEFINITION 2.10 Let the *m*-tuples

$$r_1 = \langle a_1, b_1, \ldots, m_1\rangle, \qquad r_2 = \langle a_2, b_2, \ldots, m_2\rangle$$

represent each one reaction in a given chemical system. Then the *sum* of the two reactions in the same system is the third reaction represented by

$$r = r_1 + r_2 = \langle a_1 + a_2, b_1 + b_2, \ldots, m_1 + m_2\rangle.$$

Since in principle every reaction has an inverse, any two reactions in a chemical system can add up to form a third, and the null reaction adds to any reaction leaving it unchanged, we have (Aris, 1965)

THEOREM 2.1 Let R be the set of all chemical reactions that can occur in a given chemical system, $+$ the binary operation of reaction addition, $-$ the unary operation of reaction inversion, and $\mathbb{O}$ the null reaction. Then the structure $\mathscr{R} = \langle R, +, -, \mathbb{O}\rangle$ is a group.

Now, whereas some reactions are concurrent, others are successive. The notion of reaction succession is obvious:

DEFINITION 2.11 Let $\langle R, +, -, \mathbb{O}\rangle$ be the group of possible reactions in a chemical system, and call

$$r_i = \mathbb{F}_c(t_i) - \mathbb{F}_c(t_f), \qquad r_j = \mathbb{F}_c(t_i') - \mathbb{F}_c(t_f')$$

two members of R. Then r_i *precedes* r_j in time iff r_j sets on after r_i. I.e.

$$r_i < r_j =_{df} t_i' \geq t_f.$$

This concept forms part of a notion of particular interest to the chemistry of living systems, namely that of pathway. A pathway is a set of reactions such that the product of one occurs as the reactant in the next. More precisely, we make

DEFINITION 2.12 Let $\langle R, +, -, \mathbb{O} \rangle$ be the group of possible reactions in a chemical system, and call

$$R_n = \langle r_i \in R \mid i \in \mathbb{N} \rangle$$

a list of n members of R. Then R_n is a *pathway* or *reaction sequence* iff, for any value of i occurring in R_n,

(i) r_i precedes r_{i+1} in time, i.e. $r_i < r_{i+1}$, and

(ii) some reaction products of r_i are reactants of r_{i+1}.

In closing this subsection, it will be fair to warn the reader that there are alternative ways of representing chemical compounds and chemical reactions, in particular those proposed by Aris (1965); Sellers (1967, 1970, 1971); Krohn, Langer and Rhodes (1967); and Feinberg (1972). All these alternative representations accept the usual chemical notation, which conflates individual chemicals (e.g. molecules) with their species, writing for example 'H_2O' for both the water molecule and the chemical species water. Another ambiguity of the usual chemical notation is that, when "reading" (interpreting) a reaction schema such as $\ulcorner A + B \rightharpoonup C + D \urcorner$, we are asked to interpret the first '+' as symbolizing the combination of the reactants, and the second '+' as the mixture of the reaction products. For these reasons the standard chemical notation is a paradigm of double talk – of Platonism and materialism. Therefore we had to go our own, even if mathematically less sophisticated, way.

1.5. *Chemical Control*

Every system is controlled by its environment; in particular, the milieu of a chemical system is a variable source and sink of chemicals and energy. If the environment is "fit", the chemical system will come into being and subsist for a while. On the other hand a "hostile" environment, e.g. one that supplies either too little or too much energy, or else one that is not wet enough, or far too oxydating (like Mars), will preclude certain reactions or even all of them. Therefore the existence of a chemical system is a sufficient condition for that of a "congenial" environment, i.e. one compatible with the reactions going on in the system.

Certain chemical systems have internal controls in addition to the external or environmental ones. That is, some of either the reactants or the reaction products of a chemical system contribute to controlling the pace of the reactions, often to the point of starting them or bringing them to a halt. In some cases a reaction is accelerated by the action of a catalyzer (e.g. an enzymatic protein). In particular the reaction product itself may catalyze the reaction (*autocatalysis*). In other cases the reaction product, as it builds up, inhibits the further progress of the reaction until the latter stops even if the supply of reactants is not exhausted (*product inhibition*). Or the catalyzer itself may at some point be broken down by some other enzyme – as is the case with a protein when no longer "needed". And a number of processes of these various kinds may occur sequentially (in particular cyclically) or even at the same time in the same system, e.g. a living cell (cf. Eigen, 1971; Lehninger, 1975).

We are particularly interested in chemical systems equipped with controls, because the living cells and its organelles are of that kind. Hence we define explicitly:

DEFINITION 2.13 A chemical system with composition state function $\mathbb{F}_c = \langle F_1, F_2, \ldots, F_m \rangle$(in some representation) is *self-controlled* iff some of the components of $\mathbb{F}_c$ depend exclusively upon other components of it (rather than on any components of the total $\mathbb{F}$ representing properties of the environment of the system).

Consider for example an enzymatic reaction in an aqueous solution – the standard biochemical process. Let E denote an enzyme acting upon a substrate S to form an intermediate compound ES (or enzyme-substrate complex), which eventually dissociates into a product P with liberation of E. Moreover suppose that this is a product inhibition reaction, i.e. one in which the product gets in the way of the reaction because it binds with the enzyme, thus inhibiting the latter from acting upon the substrate. In conventional chemical notation, we have

$$\begin{array}{ccccccc} \mathrm{E} + \mathrm{S} & \underset{k'_1}{\overset{k_1}{\rightleftharpoons}} & \mathrm{ES} & \underset{k'_2}{\overset{k_2}{\rightleftharpoons}} & \mathrm{EP} & \underset{k'_3}{\overset{k_3}{\rightleftharpoons}} & \mathrm{E} + \mathrm{P} \\ t_1 & & t_2 & & t_3 & & t_4 \end{array}$$

where the last two reaction constants are comparable, i.e. $k_3 \approx k'_3$. We leave the reader the task of representing the six reactions in our state space

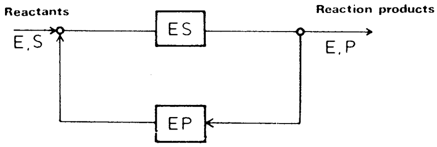


Fig. 2.5. A self-controlled chemical system of the negative feedback type. The reaction product P combines with the catalyzer E forming the 'ntermediate compound EP and thus slows down the left-to-right reactions.

notation. At this point we are only interested in emphasizing that the composition of our system is variable and that it is a self-controlled system and moreover one that is self-accelerated during certain periods and self-decelerated during others. The system can be diagrammed like an ordinary controlled system: see Figure 2.5.

Control systems of this kind fall under the general theory of control. However, chemical self-controlled systems do not accord with the now popular paradigms of control systems, namely the balancing wheel, the thermostat, and the self-controlled missile. Indeed in the latter cases the control is effected by a detachable subsystem. Moreover the whole system has fixed components and its behavior is usually assumed to be linear. Being detachable from the controlled system, a controlling system of this kind can be studied separately, not only as a component of the total self-controlled system. On the other hand if either the enzyme or the reaction product were removed from our chemical system represented by Figure 2.5, the system would cease to be chemical: it would become a physical system. Second, by definition of a chemical system, its composition is not constant but varies in the course of time. Thirdly, most of the chemical kinetics equations – in particular the ones for the formation rate of ES from the reactants – are nonlinear. These three peculiarities of chemical self-controlled systems seem to have prevented most scientists and engineers from recognizing them as systems, hence from using the systems theoretic approach to chemical problems.

2. BIOCHEMICAL SYSTEM

2.1. *Biomolecule*

We shall call *biomolecule* any macromolecule that, like a protein or a nucleic acid, is essential to some typically biological process such as cell division, morphogenesis, or development. And we shall call *biochemical* any system, natural or artificial, where biomolecules are formed or broken down. There are myriads of kinds of processes of this type, and some of them culminate in the formation of subsystems of live cells, such as ribosomes: see Figure 2.6.

The key biomolecules are the proteins and the nucleic acids: as Szent–György used to say, while the former carry on the business of life the latter are the guardians of life. Both are polymers (Definition 2.4(iii)). Whereas the building blocks (monomers) of the latter are nucleotides of mainly 8 different kinds, proteins are formed by amino acids, of which 20 species are commonly found in bioproteins. The biomolecules are then macromolecules, and they are constituted by so many units that they can attain remarkable sizes and weights. Because they are so large they are quite sensitive to environmental disturbances. Hence their stability requires fairly constant physical environmental conditions. To a first approximation some such molecules can be modeled as very long chains that can fold-in forming closely packed particles. (Cf. Vol'kenshtein, 1970; Lehninger, 1975; Watson, 1976.)

We shall borrow from biochemistry the notions of nucleotide and of amino acid. We shall also borrow the postulate that, whereas the nucleic acids are formed from eight nucleotide species, proteins involve 20 kinds of amino acids. (We shall disregard proteins other than those occurring in living beings.) More precisely, we incorporate the following biochemical axiom in our ontology:

POSTULATE 2.3. The family M of molecular species contains eight species of mononucleotides and 20 species of amino acids.

Each of the members of the mononucleotide family is composed of phosphoric acid, a sugar, and a nitrogenous base. There are two genera of nucleotides: the ribonucleotides and the deoxyribonucleotides. These genera differ by the kind of sugar: the former contain ribose whereas the latter contain deoxyribose. (A RNA molecule is a polymer of blocks of the ribo type, and a DNA molecule is a string of units of the deoxyribo kind.) In turn, there are four species of ribonucleotides and another four of deoxyribonucleotides.

The nitrogenous bases of greatest biological interest are adenine (A), guanine (G), cytosine (C), uracil (U) and thymine (T). The first four combine with ribose and phosphoric acid to form ribonucleotides or constituents of RNA:

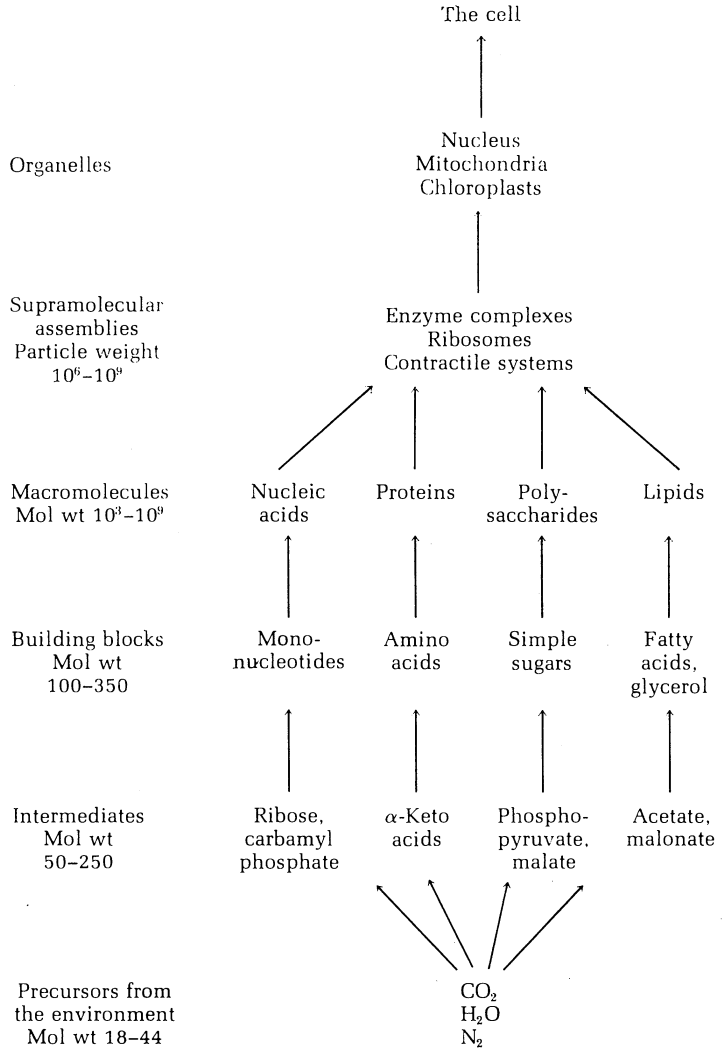

The cell
Organelles
Nucleus
Mitochondria
Chloroplasts
Supramolecular assemblies
Particle weight
10^6–10^9
Enzyme complexes
Ribosomes
Contractile systems
Macromolecules
Mol wt 10^3–10^9
Nucleic acids
Proteins
Poly-saccharides
Lipids
Building blocks
Mol wt
100–350
Mono-nucleotides
Amino acids
Simple sugars
Fatty acids, glycerol
Intermediates
Mol wt
50–250
Ribose, carbamyl phosphate
α-Keto acids
Phospho-pyruvate, malate
Acetate, malonate
Precursors from the environment
Mol wt 18–44
CO_2
H_2O
N_2

Fig. 2.6. Build-up processes that start with simple molecules in the environment and end up in living cells. 'Mol wt' is short for 'molecular weight'. From Lehninger (1975).

$$a = \text{AMP}, \qquad g = \text{GMP}, \qquad c = \text{CMP}, \qquad u = \text{UMP}.$$

('AMP' is short for 'Adenosine 5′-phosphoric acid'. Similarly with the other three.) On the other hand the nitrogenous bases A, G, C and T combine with deoxyribose and phosphoric acid to form deoxyribonucleotides or constituents of DNA:

$$a' = \text{dAMP}, \qquad g' = \text{dGMP}, \qquad c' = \text{dCMP}, \qquad t' = \text{dTMP}.$$

In short, there are two sets of units (monomers) from which the polynucleotides or nucleic acids are formed:

Ribonucleotides $R = a \bigcup g \bigcup c \bigcup u$
Deoxyribonucleotides $D = a' \bigcup g' \bigcup c' \bigcup t'$.

We proceed to characterize the nucleic acids and in the first place the members of the DNA family. We start with the building blocks in the set $D = a' \bigcup g' \bigcup c' \bigcup t'$ and form the set of all possible strings or sequences of members of D, with the sole restriction that the composition of each compound be equal to the union of the compositions of its constitutents or precursors. I.e. if x and y are in D and combine to form $z = xy$, then $\mathscr{C}_A(z) = \mathscr{C}_A(x) \bigcup \mathscr{C}_A(y)$, where '$\mathscr{C}_A(u)$' stands for the atomic composition of u. (Cf. Chap. 1, Sec.1.2.) The new set D^* formed from members of D contains all the conceivable combinations, with repetitions and permutations, of deoxyribonucleotides. In particular D^* contains all the polynucleotides forming the DNA molecules as well as monotonous strings such as $x^n = xxx...x$ with $x \in a'$. We shall restrict our attention to those molecules that contain all four nucleotides, i.e. members of a', g', c', and t'.

Call D_4^* the subset of D^* formed by the strings containing *all four* deoxyribonucleotides in all conceivable numbers and locations, such as $wxyz$, $zxwyx$, $yywxxyzzz$, where $w \in a'$, $x \in g'$, $y \in c'$, and $z \in t'$. The set D_4^* contains all the components of all the DNA molecules we need to consider – both the actual and the merely possible ones. The structure of D_4^* is that of a semigroup. More precisely, $\mathscr{D}_4 = \langle D_4^*, \cdot \rangle$ is a subsemigroup of the free semigroup $\mathscr{D} = \langle D^*, \cdot \rangle$ generated by D.

DNA molecules are double strands of things of kind D_4^*. These double strands are coupled in such a way that every *A* base in one of the strands is paired off to a *T* base in the other, and likewise every *C* base is paired off to a *G* base in the other. Hence we make

POSTULATE 2.4 Let $\mathscr{D} = \langle D^*, \cdot \rangle$ be the free semigroup generated by the set $D = a' \cup g' \cup c' \cup t'$ of deoxyribonucleotides under the operation $\cdot$ of chemical combination, subject to the condition that, for any x and y in D^*, if $z = xy$, then $\mathscr{C}_D(z) = \mathscr{C}_D(x) \cup \mathscr{C}_D(y)$. Further, call $\mathscr{D}_4 = \langle D_4^*, \cdot \rangle$ the subsemigroup formed by those sequences that contain molecules of all four kinds in D. There are systems composed of two chains of molecules in D_4^*, whose structure includes the bijection $p : D \to D$ such that for any $x_A \in a'$, $x_G \in g'$, $x_C \in c'$, and $x_T \in t'$,

$$p(x_A) = x_T,\ p(x_C) = x_G,\ p(x_G) = x_C,\ p(x_T) = x_A.$$

DEFINITION 2.14 The systems characterized by Postulate 2.4 are called *DNA molecules.*

Remark We are assuming that an arbitrary string of molecules satisfying Postulate 2.4 is not just a formal string but a chemically possible compound. The justification for this hypothesis is that there are no known restrictions on the stringing of nucleotides nor, for that matter, on carbon (or even boron) compounds in general. In short, using the notation of Sec. 1.2, $D_4^{**} = D_4^*$.

We proceed similarly with the RNA molecules, which are constituted by mononucleotides of the species a, g, c, and u:

POSTULATE 2.5 Let $\mathscr{R} = \langle R^*, \cdot \rangle$ be the free semigroup generated by the set $R = a \cup g \cup c \cup u$ of ribonucleotides, and call $\mathscr{R}_4 = \langle R_4^*, \cdot \rangle$ the subsemigroup formed by those sequences that contain molecules of all four kinds in R. Every member of R_4 is a possible *RNA molecule.*

The union of all kinds of DNA and RNA molecules is the totality of *nucleic acids*:

$$N = D_4^* \cup R_4^*.$$

The case of proteins is similar. Here the building blocks are the amino acids or rather some of them, such as glycine or NH_2CH_2COOH. More precisely, the *protein amino acids* form the genus

$$AA = \text{Ala} \cup \text{Arg} \cup \text{Asn} \cup \text{Asp} \cup \text{Cys} \cup \text{Gln} \cup \text{Glu} \cup \text{Gly} \cup \text{His} \cup \text{Ile} \cup \text{Leu} \cup \text{Lys} \cup \text{Met} \cup \text{Phe} \cup \text{Pro} \cup \text{Ser} \cup \text{Thr} \cup \text{Trp} \cup \text{Tyr} \cup \text{Val}.$$

This set has the semigroup property under the operation $\cdot$ of chemical combination. Also: the structure $\mathcal{AA} = \langle AA^*, \cdot \rangle$, where AA^* is the set of all possible strings or sequences of members of AA, is the *free semigroup generated by* AA. The members of this group are called *polypeptides* and the longer among them *proteins.* More often than not a protein contains at least one member of every species included in AA. However, there are important exceptions to this rule. Therefore we cannot restrict our attention to the subsemigroup formed by those sequences of amino acids that contain representatives of all 20 kinds of protein amino acids.

We summarize the foregoing in

POSTULATE 2.6 Let $\mathcal{AA} = \langle AA^*, \cdot \rangle$ be the semigroup generated by the set AA of protein amino acids. Every member of AA^* is a possible *polypeptide.*

DEFINITION 2.15 A *protein* is a polypeptide composed of at least 100 amino acids. Symbol for the set of proteins: $\mathcal{P}$.

It is estimated that there are at least 10,000 kinds of proteins in the cells of contemporary organisms. Roughly half of them are enzymes, i.e. reaction controllers.

Postulates 2.3 to 2.6 entail

COROLLARY 2.1 (i) Differences among kinds of DNA molecules consist in differences in composition or in the order (sequence) of the deoxyribonucleotides (members of D);

(ii) differences among kinds of RNA molecules consist in differences in composition or in the order (sequence) of the ribonucleotides (members of R);

(iii) differences among kinds of protein molecules consist in differences in composition or in the order (sequence) of the 20 protein amino acids (members of AA).

Molecules with the same composition but a different structure – in particular polymers differing only in the relative position of the component monomers along the chain – are called *isomers.* A polymer constituted by n distinct monomers has $1.2. \ldots n$ isomers.

So much for the composition and structure of nucleic acids and proteins. Let us now turn to their formation and breakdown.

2.2. *DNA Replication, and RNA and Protein Synthesis*

Not only are the nucleic acids the stuff of heredity: they are assumed to control whatever goes on in the cell, from protein synthesis to cell division. We must therefore take a closer look at them, particularly at their mode of emergence and the way they control the synthesis of proteins. However, we are forced to simplify, even to the point of disregarding the differences among the three kinds of RNA (ribosomal RNA, mRNA and tRNA).

The typical DNA molecule is composed by two polynucleotide chains that twist about each other forming what has become the most popular microsystem – the double helix. This has the well known property of self-duplication. This process involves the gradual detachment of the two strands, each of which serves as a mold or template for the formation of a complementary strand out of the molecules that pullulate in the rich cellular background. The attachment of these molecules is activated by enzymes. The overall process is captured by

POSTULATE 2.7 The duplication of a DNA molecule is a process whereby the molecule joins with molecules from its D-environment E to form two DNA molecules. In other words, the structure of a DNA molecule contains a function

$$\rho\colon D_4^* \times 2^E \longrightarrow D_4^* \times D_4^*$$

such that $\rho(d, e) = \langle d', d''\rangle$, with $d, d', d'' \in D_4^*$, $e \in 2^E$, and $\mathscr{C}_D(d)$, $\mathscr{C}_D(d')$, $\mathscr{C}_D(d'') \in 2^E$.

As mentioned above, DNA molecules control their own reproduction, i.e. are their own templates. But they are not the only molecular systems acting as templates: RNA molecules template for the synthesis of proteins. We must therefore elucidate the notion of a template. See Figure 2.7.

DEFINITION 2.16 Let $\tau \in T$ be a concrete system of kind T and let E be a set of things in the environment of τ. Assume that τ pre-exists a second system σ of kind Σ. Further, call i_T the identity function on T, and $\mathscr{C}_A\colon \Sigma \to 2^E$ the atomic composition function that assigns each (assembled) system σ its atomic composition $\mathscr{C}_A(\sigma) = e \in 2^E$. Then f is a *molding function* for σ, and τ a *template* for σ (or a *σ-template*) iff

(i) $f = i_T \times \mathscr{C}_A\colon T \times \Sigma \to T \times 2^E$ such that

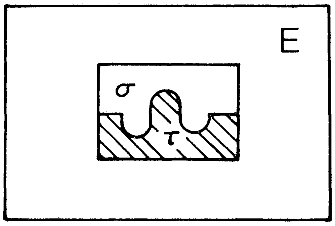


Fig. 2.7. System τ is a σ-template, i.e. τ provides a sort of mold for the assembly (e.g. synthesis) of σ out of environmental entities.

$$f(\tau, \sigma) = \langle \tau, e \rangle \quad \text{for } \tau \in T, \sigma \in \Sigma, \mathscr{C}_A(\sigma) = e \in 2^E;$$

(ii) the boundaries of τ and σ match in shape.

Molecular templates are, of course, molecules with the capacity to act as templates for the formation of further molecules. DNA and RNA are among them: DNA are autotemplates as well as RNA-templates, and RNA are in turn protein-templates. This is part of the well confirmed hypothesis misnamed *central dogma* of molecular biology, summarized in the heuristic diagrams of Figure 2.8, where the solid arrows represent the "information transfer" occurring in all cells, whereas the dotted arrows represent the "flow of sequence information" occurring in some cells.

We assume then, to begin with,

POSTULATE 2.8 RNA synthesis is the process whereby a DNA molecule

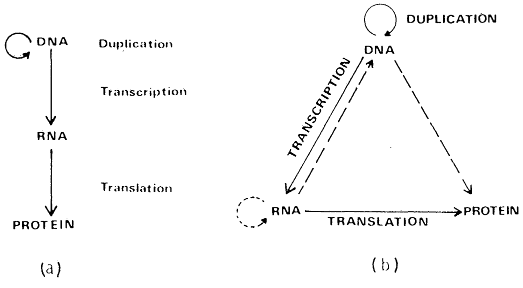


Fig. 2.8. Two versions of the "central dogma". (*a*) The standard version (e.g. Watson, 1976). (*b*) An alternative version (Crick, 1970). For a translation of the metaphorical expression 'information flow' into the nonmetaphorical language of quantum chemistry, wait a few decades.

(template) joins with nucleotides from its environment resulting in the same molecule (template) together with an RNA molecule. I.e. the structure of a DNA molecule contains a function

$$s_R = i_{D_4^*} \times \mathscr{C}_A: D_4^* \times R_4^* \longrightarrow D_4^* \times 2^E,$$

where $i_{D_4^*}$ is the identity function on D_4^* ,and $\mathscr{C}_A$ the composition function $\mathscr{C}_A: R_4^* \to 2^E$, such that

$$s_R(d, r) = \langle d, e \rangle \quad \text{for } d \in D_4^*, r \in R_4^* \; \mathscr{C}_A(r) = e \in 2^E.$$

DNA duplication and RNA synthesis are of course chemical processes, for in both cases a batch of reactants (the set e of precursors "swimming" in the environmental solution) combine into the reaction product. Moreover the entire system is a self-controlled chemosystem, since both processes are controlled by DNA molecules. However, the DNA molecules act not as ordinary chemical controls (e.g. enzymes) but as templates: theirs is template chemistry. (The same holds for the control exerted by RNA molecules in protein synthesis.) This is why the processes summarized by Postulates 2.7 and 2.8 are not represented by ordinary reaction equations, and why we had to introduce the functions ρ (DNA replication) and s_R (RNA synthesis). But this is precisely one of the peculiarities of biochemosystems in contrast to plain chemosystems.

Another peculiarity of biochemosystems is their capacity to synthesize biomolecules and, in particular, proteins. As will be recalled from Definition 2.14, a protein is a compound of amino acids and, more particularly, a member of the set AA^* of strings or sequences containing (protein) amino acids of all kinds. Again, the synthesis of proteins is controlled by RNA molecules acting as templates. More precisely, we introduce

POSTULATE 2.9 Protein synthesis is a process whereby an RNA molecule (template) joins with amino acids from its environment E, to yield the same molecule together with one protein. I.e. the structure of RNA contains the function

$$s_P = i_{R_4^*} \times \mathscr{C}_A: R_4^* \times AA^* \longrightarrow R_4^* \times 2^E,$$

where $\mathscr{C}_A: AA^* \to 2^E$, such that

$$s_P(r, p) = \langle r, e \rangle,$$

where $r \in R_4^*$, $p \in \mathscr{P} \subset AA^*$, $\mathscr{C}_A(p) = e \in 2^E$.

This axiom describes only the overall process: it does not tell us which segments of the RNA molecule correspond to which amino acids in the sequence characterizing the given protein. The fine structure – though by no means the mechanism – of this matching is given by the misnamed *genetic code*. This "code" is nothing but the detailed correspondence between the nucleotides in the RNA template and the amino acids in the protein synthesized on it. Let us describe this correspondence.

Let $R = a \cup g \cup c \cup u$ be the set of ribonucleotides forming the RNA molecules in R^*_4 (recall Postulate 2.5). The latter is the set of strings (or sequences) containing ribonucleotides of each of the four nitrogenous bases, and therefore all the possible triplets of such, e.g. *uca*, *uuc*, and *gcg*, enter into the composition of members of R^*_4. Call R^*_3 the set of $4^3 = 64$ such triplets, called *codons*. See Figure 2.9.

Take next the set AA of 20 kinds of protein amino acids, i.e. the protein building blocks. The full genetic "code" is the function that assigns each codon (triplet of nucleotides in an RNA molecule) a given amino acid in a protein. (Cf. Watson, 1976.)

We reformulate the preceding as a convention and an axiom:

DEFINITION 2.17 Let R^*_3 be the set of 64 triplets of nucleotides in $R = a \cup g \cup c \cup u$ such that, for every $x \in R^*_3$, there exists a segment $y \in R^*_4$ of an RNA molecule for which $x \sqsubset y$ (x is a part of y). Every member of R^*_3 is a *codon*.

POSTULATE 2.10 The codon-amino acid (or RNA-protein) correspondence is the partial function $\pi\colon R^*_3 \to AA$ that

(i) assigns every codon, other than the starting and the terminating codons, one amino acid in a protein. I.e. with the noted exceptions, if $r \in R^*_3$, then $\pi(r) = a \in AA$, where $a \sqsubset p$ and $p \in \mathscr{P}$ is some protein;

(ii) is order preserving: i.e. if $r, r' \in R^*_3$, and r precedes r' in a given codon sequence, then $\pi(r)$ precedes $\pi(r')$.

In the literature a different function f occurs implicitly, namely one assigning each amino acid one codon, i.e. $f\colon AA \to R^*_3$, where $'f(a) = r'$, for

RNA strand *aaugcuuaaaccgggauu*...
Triplets (*aau*)(*gcu*)(*uaa*)(*acc*)(*ggg*)(*auu*)...

Fig. 2.9. The sequence of nucleotides in the RNA molecule, when acting as a template for protein synthesis, is organized into triplets (codons). The kind and the position of each amino acid in the protein are determined by the composition and the position of the corresponding codon.

$a \in AA$ and $r \in R_3^*$, is read '*r codes for a*'. However, it has been found experimentally that different codons can "code for" one and the same amino acid. (I.e. the "code" is degenerate or many-one, hence not a proper code allowing one to break any ciphered molecular message.) In other words, no such function f from AA to R_3^* exists, so that the converse "translation" of protein structure into RNA structures is impossible.

A genuine code is a certain correspondence between two sets of *artificial* signs and, more particularly, *languages* (cf. Birkhoff and Bartee, Ch. 8). The coding and decoding functions of errorless codes are one to one, not many to one like π. Therefore the use of the linguistic concept of code in chemistry is metaphorical. The same holds for the expressions 'the templates instruct (or code for) protein synthesis' and 'the genetic instructions contained in a gene'. That these are analogies is shown by the following facts. First, a genuine information system is one composed of a transmitter, a receiver, and an information channel between them; no such components are apparent in a chemical system. Second, the computations of the quantity of information allegedly locked in a chunk of genetic material (i.e. a segment of a DNA molecule) are phoney – so much so that they occur in no biological law statements and everyone gives his own favorite arbitrary estimate. For these reasons the information-theoretic language employed in molecular biology is a heuristic prop that has now exhausted its heuristic power. Moreover it is a stumbling block on the road to transforming the "central dogma" into a hypothetical-deductive system explaining the synthesis and the duplication of nucleic acids, as well as the synthesis of proteins, exclusively in physical and chemical terms instead of describing them in analogical terms such as 'informational molecule', 'information flow', 'transcription', and 'translation'.

Note finally that, like the previous mappings introduced in this section – namely ρ (DNA replication), s_R (RNA synthesis) and s_P (protein synthesis) – π is not defined by Postulate 2.10, which characterizes a whole class of such functions. However, unlike ρ, s_R and s_P, π is known in detail: indeed it is given by the table that summarizes the so-called genetic code. What is still unknown is the origin of the "code". Nor is it known whether nucleic acids have always been "used" to synthesize proteins.

2.3. *Biochemical System*

A biochemical system may be characterized as a system wherein biomolecules are synthesized or broken down. A ribosome and an insulin synthesis reactor are biochemical systems. On the other hand a cellular membrane is a plain chemosystem because it is not involved in the production or disintegration of biomolecules.

Our explicit characterization is

DEFINITION 2.18 A system σ is a *biochemical system* (or a *biochemosystem*) iff σ is a self-controlled chemical system such that

(i) the composition of σ includes proteins or nucleic acids;

(ii) the environment of σ contains all the precursors of the components of σ and moreover is compatible with the processes of synthesis of proteins or of nucleic acids;

(iii) the components of σ exchange things and energy with the environment (i.e. σ is semi-open) and moreover some of them participate in the reactions happening in σ;

(iv) the structure of σ contains the protein synthesis (or the breakdown), the DNA replication, or the RNA synthesis function.

The reactions in a chemical system can eventually come to a halt because of exhaustion of reactants or as a result of product inhibition. In contrast, a biochemosystem is capable – environment willing – of maintaining its reactions or switching over to alternative reactions by predating on its environment and self-assembling its catalyzers (enzymes), and eventually also its templates. In fact, processes of two kinds are constantly going on and intertwining in a biochemical system: (*a*) the building up or self-assembly of complex molecules from simpler precursors (*anabolism*), and (*b*) the breaking down or dismantling of complex chemicals (*catabolism*). Furthermore these two processes are interdependent: the break down processes require enzymes that had to be synthesized to begin with, and such syntheses are often dependent upon further enzymes occurring as intermediates in breakdown processes. Moreover the anabolic processes consume energy released chiefly by catabolic ones. In short, biochemical systems are characteristically metabolic. We refine this notion by introducing

DEFINITION 2.19 If σ is a biochemical system, then

(i) the *intermediary metabolism* of σ is the set of all the enzymatic reactions involving only components of σ;

(ii) the *metabolism* of σ is the set of all the chemical reactions involving components of σ and the environment of σ.

Certain cell components are metabolizing biochemosystems. However, it is assumed that the precursors of the first organisms, on the Earth and probably elsewhere, were extracellular biochemical systems, and that the first cells emerged by self-assembly from, or the merger of, some such biochemical systems. And it is sometimes conjectured that the immediate precursors of cells may have been systems containing both proteins and nucleic acids. Such systems, though not possessing all the properties of contemporary cells, would have their main emergent properties, namely the abilities to metabolize, self-reproduce, and mutate. (Cf. Küppers, 1975.) However, the problem of the origin of life will be taken up in the next chapter.

2.4. *Concluding Remarks*

In this chapter we have distinguished three kinds of system that are usually conflated: physical, chemical, and biochemical. See Figure 2.10.

Chemical systems are distinguished from plain physical systems by the fact that some of their components are continually engaged in reactions in the course of which some chemicals are consumed while others are produced. Chemical systems are thus paradigms of qualitatively changing things. Not that qualitative change fails to occur in physical systems: the emission and absorption of light are qualitative changes, and so are nuclear fission and fusion as well as changes in the state of aggregation. However, many a physical system can be rendered immune to qualitative changes for a while. In contrast, a chemical system is by definition one whose composition changes relentlessly, so that it eventually becomes

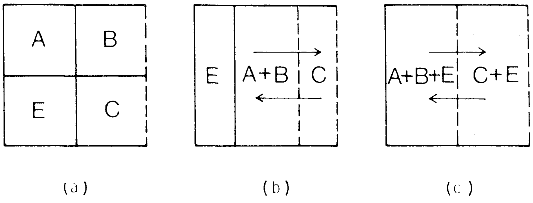


Fig. 2.10. Three kinds of system: (*a*) physical system with four non-reacting components; (*b*) chemical system with one nonreactive component (E); (*c*) biochemical system: all components react, and one of them (E) is an enzyme or proteinic catalyzer. Whereas physical and chemical systems may be almost completely closed (full line) or open only in some respects, biochemical systems must be semiopen (dashed line), for they prey on their environment.

something qualitatively different from what it was in the beginning. That such changes can be explained, at least in principle, with the help of physical theories – though not with them alone – does not turn them into physical processes. We are here concerned with ontological, not epistemological matters: with objective level distinctions not with the understanding of the higher levels with the help of knowledge concerning the lower ones.

Biochemical systems are the seats of even deeper and more sustained qualitative changes. They engage not just in any old chemical reactions but in reactions resulting in the self-assembly or the self-dismantling of huge molecules (often millions of times bigger than the inorganic ones). Moreover some such reactions involve templates, objects that do not occur in ordinary chemical reactions. Finally biochemical systems have a sustained activity thanks to their enzymes, and they can help themselves to environmental morsels. Because of these properties biochemical systems come rather close to biosystems.

The differences among physical, chemical, and biochemical systems are not just differences in complexity: they are momentous qualitative differences, so great in fact that we are justified in speaking of different levels: the physical and chemical levels, and perhaps also the biochemical level. Of course we must not forget that all biochemical systems are chemical systems, which in turn are composed of physical systems: otherwise we won't succeed in explaining biochemistry with the help of chemistry, or chemistry with the help of physics. Nevertheless we must not exaggerate similarities at the expense of differences, particularly if we hope to understand one more qualitative jump, namely the one from biochemical systems to biosystems. Let us take this leap.

CHAPTER 3

LIFE

Life has long been a mystery on which mystery mongers have thrived. It is no longer such: biologists are beginning to understand what makes organisms tick, and physicians to manipulate the control systems responsible for that ticking – which, incidentally, is not like a clock's. This understanding could not be forthcoming as long as organisms were studied either on their own level exclusively (holistic approach) or as physical systems devoid of emergent properties (levelism), and in either case apart from their history. The theory of evolution and molecular biology are to be credited for having transformed the mystery of life into the problem of life: its origin and maintenance, its evolution and extinction.

We have come to understand that many of the peculiarities of organisms have molecular roots. Thus cell division, a typically biological event, is the outcome of a chemical process. Moreover we understand now that other biological properties and processes – such as those of emerging by way of self-assembly of nonliving units, and of being subjected to natural selection – are properties shared by all systems. (Recall Ch. 1, Sec. 3.2.) In other words, we have come to realize that organisms are systems composed of chemical systems, and moreover that they have evolved from lower level things. This suggests the plan of this chapter: we start out with biochemical systems (studied in Ch. 2), endow them with certain emergent properties to obtain biosystems, and examine some peculiarities of biological adaptation and evolution.

The upshot of our study will be a merger of ontological pluralism (or emergentism) with moderate epistemological reductionism. Indeed we shall take the view that ontological monism – whether mechanistic or spiritualistic – is false because it denies the emergence of novel properties and laws as one moves from physical systems through chemosystems to biosystems. On the other hand the program of explaining the higher (more complex and newer) levels by the lower ones – i.e. epistemological reductionism – has proved fertile and is in keeping with contemporary biology. Indeed the latter – particularly molecular biology and evolutionary biology – is explaining life without explaining it away, i.e. without leveling down (let alone up), and is doing it with the help of chemistry, though not in terms of it exclusively.

1. FROM CHEMISM TO LIFE

1.1. *Self-Organization of Biochemical Systems*

According to contemporary biology (*a*) modern organisms descend ultimately from primitive biosystems (protocells), and (*b*) the latter in turn emerged from biochemical systems, which evolved out of biomolecules. (See Oparin (1968), Miller and Orgel (1974), and Oró, Miller, Ponnamperuma and Young (Eds.) (1974).) In turn, all of the biomolecules are carbon compounds, all of which contain also hydrogen and oxygen, and most of them also nitrogen. This is no accident: of all the 100-odd elements, hydrogen, carbon, nitrogen and oxygen constitute the most stable molecules and by far the greatest variety of them. Boron too has the capacity to form large stable molecules, but it is far less common than the other four elements, which exist almost everywhere in the explored part of the universe.

Moreover, the physical conditions necessary for the synthesis of biomolecules – such as plenty of solar energy and humidity, as well as moderate temperatures and pressures – have existed on our planet for about four billion years. Besides, it is reasonable to conjecture that the primitive atmosphere was reducing rather than oxydizing, hence lenient rather than hostile to the self-assembly of biomolecules and systems composed thereof. Under such conditions it is no wonder that biomolecules should have formed and eventually merged to constitute larger systems. In short, ours is "the best of all possible environments for life" (Henderson, 1913, p. 273) – or at least one among a number of environments fit for the origin and maintenance of life. This is not to say that the emergence of life was inevitable: scientists maintain only that it was possible (i.e. lawful rather than miraculous). Moreover, since the conditions for the emergence of life are rather stringent, they are also probably rather exceptional, so that the likelihood that they have occurred elsewhere in exactly the same manner must be vanishingly small. Therefore, although it is very likely that there are extraterrestrial organisms, it is unlikely that they, or at least the more complex among them, much resemble those we know (Jacob, 1977).

Until recently it was assumed to have taken aeons for biochemical systems to form from hydrogen, carbon, oxygen, and nitrogen. And this because it was taken for granted that such self-assemblies could occur only by the simultaneous chance encounters of all the atoms composing a biochemical system. Indeed, such events are so unlikely that they can be ruled out. Although chance is ever present, particularly in chemical dissociations, the chemical reactions of the buildup and substitution kinds are brought about by chemical bonds against the disorganizing random motions of the atoms and molecules in the surroundings. (Surely there is randomness in the very collisions ensuing in chemical compounds, but the latter are unstable, hence short lived, unless the atoms are held together by interatomic bonds. So, chance within the system works against chance in the environment. And even the dissociations caused by random thermal motions may result in reaction products that are in turn reactants forming eventually more stable systems.)

Nowadays we know that chemical and biochemical systems cannot help but self-assemble under the action of bonds of various kinds. Moreover we know that such self-assembly processes are more likely to occur in stages than at one stroke. In particular, the formation of primitive proteins may have proceeded in two stages: amino acid synthesis followed by polymerization. Therefore the first biochemical systems, and even the first organisms, may have formed on our planet and elsewhere as soon as the requisite conditions were met. This explains the short time span between the origin of rocks and the appearance of the first bacteria and blue-green algae.

As soon as biomolecules assembled they could associate with others forming larger molecules or transmolecular systems. About the simplest systems of the latter kind are the colloidal particles. For example, a protein molecule in an aqueous solution attracts a number of water molecules forming a colloidal hydrophilic complex. And in turn a number of such systems can coalesce to form a coacervate: see Figure 3.1. In either case the resulting system is a distinct thing, separate from its liquid medium and with a structure of its own (Oparin, 1968). Presumably nucleic acids can form similar systems. In particular a DNA molecule placed in a suit-

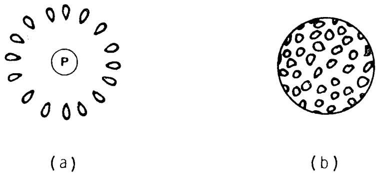


Fig. 3.1. The self-assembly of colloidal particles around a protein molecule (P) in an aqueous solution (*a*), and the coalescence of colloidal particles into a coacervate (*b*).

able solution of nucleotides and nucleic acids could perhaps start synthesizing RNA molecules, which in turn would control the synthesis of proteins as well as of further DNA molecules. Furthermore a protecting double layer of lipid molecules could self-assemble around such a DNA-RNA-protein system. Such a semipermeable membrane would increase the concentration of biomolecules, hence the rate of biochemical reactions, and would keep out disrupting (e.g. dissociating) agents. New reactions would thus be possible, ending up in the formation of new kinds of biopolymers and even of transmolecular systems such as vesicles and organelles of various kinds.

Given biomolecules in a suitable medium, the formation of organelles such as chloroplasts and ribosomes is far from miraculous. Indeed a ribosome, for example, is not an impossibly complicated system: it consists of 58 macromolecules, three of them RNA molecules, and the rest proteins of various kinds. What the ribosome does – its function – is, chiefly, to synthesize proteins from precursors in its environment.

The next step is, of course, the self-assembly of a few such transmolecular systems to form a cell, however primitive – hence inefficient, ergo ill-adapted. The rest is history – evolutionary history driven by mutation and selection. In short, the process of emergence of life may have looked like Figure 3.2.

This diagram condenses what is common to various rival hypotheses concerning the precise self-assembly of primitive cells. (See Florkin (1960), Oparin (1968), Fox and Dose (1972), Miller and Orgel, (1974), Oró, Miller, Ponnamperuma and Young (Eds.) (1974).) There is rapidly accumulating

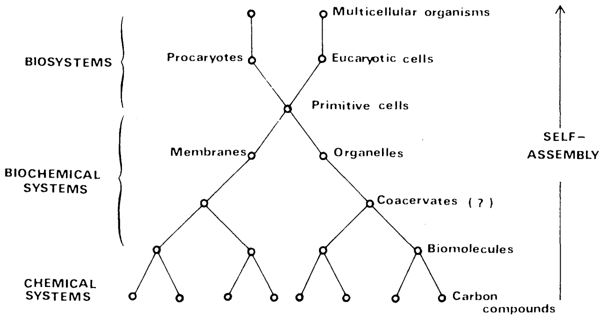

Multicellular organisms
BIOSYSTEMS
Procaryotes
Eucaryotic cells
Primitive cells
SELF–ASSEMBLY
Membranes
Organelles
BIOCHEMICAL SYSTEMS
Coacervates (?)
Biomolecules
CHEMICAL SYSTEMS
Carbon compounds

Fig. 3.2. The step-wise self-assembly of organisms from carbon compounds.

evidence that most of the buildup processes depicted in Figure 3.2 have been going on on our planet, and at least the first few steps elsewhere as well. In particular, the analysis of Precambrian rocks, of meteorites, and recently of some celestial objects, has been yielding amino acids, fatty acids, and even nucleic acids.

We compress and generalize the above into

POSTULATE 3.1 Let $\Sigma = \langle \Sigma_i \mid 1 \leq i \leq n \rangle$ be a sequence of really (nomologically) possible species of system, such that the systems of kind Σ_i are composed of systems of kind Σ_{i-1}. Then every actual system belongs in the sequence Σ and has self-assembled from systems in the immediately preceding term of the sequence.

1.2. *Definition of a Biosystem*

We have reached the point where we can list the properties of a system we believe to be jointly necessary and sufficient for it to qualify as a living thing. We assume that an organism is a system such that

(i) its *composition* includes proteins (both structural and functional, in particular enzymatic, the latter enabling it to exploit its environment) as well as nucleic acids (which make for its reproducibility and the likeness of its offspring);

(ii) its *environment* includes the precursors of all its components (and thus enables the system to self-assemble most if not all of its biomolecules);

(iii) its *structure* includes the abilities to metabolize, to self-repair, and to reproduce.

We shall spell out this hypothesis into an axiom and a convention. Here is the former:

POSTULATE 3.2 There are systems of a kind B such that, for every member b of B,

(i) b is composed of chemical and biochemical (hence metabolizing) subsystems;

(ii) b incorporates some of the biomolecules it synthesizes (rather than releasing them to its environment);

(iii) the possible activities of b include (*a*) the rearrangement, assembly, and dismantling of components; (*b*) the replacement of nonfunctional components and the removal or neutralization of some chemicals; (*c*) the capture and storing of free energy (in ATP molecules, glycogen, and fat) for future consumption;

(iv) *b* can adjust to some environmental changes;

(v) *b* is a component of some supersystem (organism) or else of some population of systems of the same kind;

(vi) some of the subsystems of *b* are capable of reproducing parts of *b*;

(vii) some of the subsystems of *b* control some of the processes occurring in *b* in such a way that a fairly constant inner environment (*milieu intérieur*) is maintained in the system;

(viii) all of the control systems in *b* are interconnected by signals (diffusion of chemicals, propagating chemical reactions, electrical signals, etc.) and so constitute a signal or information network;

(ix) one of the control systems of *b* (its genetic system or genome) is composed of nucleic acid molecules, is unique to *b*, and controls the development and the reproduction of *b*;

(x) the descendants of *b* are similar to *b* but may possess some new traits of their own (by mutation or gene recombination);

(xi) *b* competes in some respects with other systems of the same species as well as of other species, and cooperates with the same or others in some other respects;

(xii) *b* descends proximately or remotely from biochemical systems devoid of some of the properties listed above, and its nucleic acid molecules keep traces (hence some record) of the ancestry of *b*;

(xiii) *b* lasts, as a member of *B*, for a limited time lapse.

The name of the *B*'s should come as no surprise:

DEFINITION 3.1 The systems *B* referred to by Postulate 3.2 are called *biosystems* or *living things.*

Surely biosystems have many more properties than those we have just listed, but these seem to be necessary and sufficient to distinguish terrestrial biosystems from other things. In particular, clause (ii) exhibits one of the pecularities of biosystems as opposed to nonliving systems. In fact it is possible to assemble in the laboratory systems capable of synthesizing biomolecules that do not go into forming cell components such as organelles and membranes, let alone whole cells. To put it in metaphorical and therefore unscientific terms: a characteristic of metabolism in biosystems is that, far from being either "indifferent" or "self-serving", it "serves" the organism as a whole. (No teleonomy here: those systems that do not metabolize this way do not make the grade: they are either nonliving or sick.) Likewise clause (iii): a nonliving (e.g. biochemical) system need not be self-cleaning and self-repairing but could accumulate reaction-inhibiting chemicals (e.g. toxic substances) that would eventually bring some or even all of its reactions to a halt.

Note that in clause (xi) we have included not only competition, the heart of Darwinism, but also cooperation. Of course in the vast majority of cases cooperation is not deliberate, i.e. it does not consist in solidarity, but it is not less real for that. Without the unwitting cooperation of plants, animals, and other organisms, the terrestrial atmosphere might not be very different from what it was in the beginning, i.e. about five billion years ago: for one thing it might still be reducing rather than oxydizing. Note also a peculiarity of heredity in biosystems (clause (xii)). Like other complex systems, such as stars, mountains, and artifacts, biosystems betray their past. But, unlike nonliving systems, biosystems concentrate their archives in a tiny part of themselves, namely their genome.

Not all of the functions (properties, processes and activities) we attribute to an organism are actually carried out by it throughout its life history. Thus metabolism and cellular division may be temporarily suspended, as is the case with spores and dormant seeds. Some of the properties of biosystems are then dispositions, or potential properties that may actualize under favorable environmental circumstances. (For the notion of potentiality see Vol. 3, Ch. 4.)

We have not included certain properties that are obviously possessed by biosystems – and therefore often listed as definitory – because they are also common to all biochemical systems and even to some physical systems. Among them are (*a*) having a more or less definite boundary separating the system from its environment, and (*b*) being subject to natural selection. That the former property is shared by many a macrophysical nonliving system, is clear. As for the latter property, Postulate 1.6 attributes it to all systems.

Finally note that, according to Definition 3.1, chromosomes are not alive because they do not metabolize, and ribosomes are not alive because they do not replicate. Likewise viri are not alive because they do not function at all outside some host cell – so much so that independent viri are often crystals. (Only the host cell-virus system is alive.) Nor do robots, however sophisticated, qualify as biosystems, if only because they are made of mechanical and electrical components instead of biochemical ones. A couple of properties, no matter how important, do not suffice to characterize a biosystem; the property system consisting of the 13 properties listed in Postulate 3.1 is necessary and sufficient – until further notice.

1.3. *Cell and Organism, Biospecies and Biopopulation*

To begin with let us take a closer look at a feature of biosystems that is of particular philosophical interest, namely their cellular composition. To identify the smallest unit of living matter, as well as that of life, we shall use the notions of biosystem (Definition 3.1) and of system component (Definition 1.2):

DEFINITION 3.2 (i) A *cell* is any biosystem such that none of its components is a biosystem. In obvious symbols,

$$x \in C =_{df} x \in B \;\&\; \forall y\,(y \in \mathscr{C}(x) \Rightarrow y \notin B);$$

(ii) the *cellular composition* of a biosystem is the set of cells that compose it:

$$x \in B \Rightarrow \mathscr{C}_C(x) =_{df} \mathscr{C}(x) \cap C, \quad \text{where} \quad \mathscr{C}(x) = \{y \in \Theta \mid y \sqsubset x\};$$

(iii) an *organism* is a cell or a multicellular biosystem which is not a proper subsystem of a biosystem:

$$x \in O =_{df} x \in C \vee x \in B \;\&\; (\mathscr{C}_C(x) \subset C \;\&\; \neg(\exists y) \quad (y \in B \,\&\, x \prec y)$$

Remark 1 The intersection of *C* and *O* is the set of unicellular organisms. *Remark 2* Useful synonyms for *C* and *O* are *cellular level* and *organismic level* respectively. *Remark 3* According to the above definition, the living part of a coral reef may be regarded as an organism. *Remark 4* A bacterial culture, though a system, is not a biosystem; in fact it is a population. *Remark 5* It has been held that organisms could be construed either as "classes of cells' or as "parts of biological species" (Hull, 1974, p. 48). Both disjuncts are false. Classes, in particular biospecies, are sets, hence concepts, whereas organisms are things and, more particularly, concrete systems. A thing can be a member (∈) of a certain set and a part (⊏) of another thing: it cannot be a part of a set and a member of a thing, and this by definition of the membership and the part-whole relations.

We can now define the concept of a biospecies. We shall construct it with the help of the notion of a natural kind (Vol. 3, Ch. 3, Sec. 3.5) as determined by a bunch of natural laws, and the concept of descent (Definition 1.17):

DEFINITION 3.3 A species is a *biospecies* iff

(i) it is a natural kind (rather than an arbitrary collection);

(ii) all of its members are organisms (present, past, or future);

(iii) it descends from some other natural kind (biotic or prebiotic).

Remark 1 This definition avoids the difficulties that beset some of the hazy notions of a biospecies encountered in the biological and biophilosophical literature. In particular we are not requiring reproductive isolation, for this is irrelevant to organisms that do not reproduce sexually, and it founders in the cases of plant and animal hybrids that fail to be sterile. *Remark 2* We do not confuse a biospecies with a biosystem, as even some eminent biologists have done. (E.g. Dobzhansky, 1970, p. 354: "A species is composed of individuals as an individual is composed of cells, or as a termite or ant colony is composed of fertile and sterile members.") Like any other species, a biospecies is a set, hence a concept. Only, it is not an arbitrary set but one defined by a definite collection of laws. And a biospecies is not an unstructured set but one partially ordered by the relation of descent. But a species is not a concrete system, because (*a*) its members do not all share the same environment, and (*b*) its past and future members are not coupled to its present members. *Remark 3* Because of clause (iii) of Definition 3.3, the organisms that may eventually be synthesized in the laboratory would belong not to biospecies but rather to artificial biospecies.

Finally let us deal with three bioentities that are neither individual organisms nor sets of such:

DEFINITION 3.4 A system is

(i) a *biopopulation* iff it is composed of individuals of the same biospecies;

(ii) a *community* or *ecosystem* iff it is composed of several interacting populations of organisms belonging to different biospecies;

(iii) a *biosphere* iff it is the largest system in which a given biosystem participates.

Remark 1 It is usual to include, in the definitions of biopopulation and ecosystem, the sharing of a territory or habitat, such as a swamp or a tree canopy. This is unnecessary in our case because we have defined those entities as concrete systems: they would not be such unless there were bonds amongst its components, and such bonds are possible only if the individuals concerned are not far apart. *Remark 2* Nor is it necessary to specify the planet in defining the general notion of a biosphere: the definition applies to any planet and it does not hold for the set of ecosystems on all possible planets because such a set is not a concrete system.

1.4. *Biolevel*

We have distinguished, more or less explicitly, six different biotic levels, namely those of cell, living component of an organism, organism, population, ecosystem, and biosphere. We may as well display them for reference purposes:

B_1 = *cell level* = the set of all cells
B_2 = *organ level* = the set of all organs
B_3 = *organismic level* = the set of all organisms
B_4 = *population level* = the set of all populations
B_5 = *ecosystem level* = the set of all ecosystems
B_6 = *biosphere level* = the set of all biospheres.

The family $\beta = \{B_1, B_2, B_3, B_4, B_5, B_6\}$ illustrates the concept of a level structure (Definition 1.8). So, the set β together with the level precedence relation deserves a name of its own:

DEFINITION 3.4 The set β of biotic levels together with the relation $<$ of level precedence, i.e. $\mathscr{B} = \langle \beta, < \rangle$, is the *biolevel structure*.

$\mathscr{B}$ has also been called the *scala naturae* and the *hierarchy* of life. These are misnomers: a ladder proper is supposed to lead somewhere, and a hierarchy proper involves a dominance or subordination relation, while $\mathscr{B}$ does not.

Now, the biotic levels we have included in β are not the only logically possible ones. In particular, one can imagine the existence of subcellular organisms; in fact it has occasionally been claimed that such biosystems exist. We shall assume that whatever is alive or composed of living beings is in β:

POSTULATE 3.3 Every biosystem and every system composed of biosystems belongs in some level of the biolevel structure $\mathscr{B} = \langle \beta, < \rangle$.

This version of the "hierarchical organization" of life is static: it talks about levels and their order but not about their origin. However, most biologists nowadays would claim that biotic levels are so many stages of an evolutionary process: that every biolevel has emerged spontaneously from the preceding (prebiotic or biotic) level. Actually we need no separate axiom to incorporate and refine this assumption, as it is a mere specification of Postulate 3.1:

COROLLARY 3.1 Every concrete system belonging to a given biolevel has self-assembled from things at the preceding level.

In turn, an immediate consequence of this corollary is that every system at a given biolevel is preceded in time by its components or precursors, Hence level precedence and temporal precedence, though not cointensive, are coextensive. The biolevel structure is no longer static: it has become part of an ontology which is not only pluralistic but also evolutionary. Biolevels are not static layers that happen to be piled atop one another. They succeed each other in time (metaphorically speaking) and they do so by virtue of a definite and pervasive mechanism, namely self-assembly. And they are not rungs in a hierarchy leading from lowly atom through middling man to the Supreme Being: levels are but stages in a natural evolutionary process that may have occurred and indeed may be occurring at various places and epochs in the history of the universe, though not twice in the same manner.

1.5. *Concluding Remarks*

We have characterized biosystems as systems composed of chemical and biochemical subsystems. Hence while some of the properties of biosystems can be traced back to their biochemical constituents, others are emergent or peculiar to the former. The most obvious property of this kind is of course that of being alive – a derivative property not a root one. Thus the chloroplasts and genes of a plant cell are essential components of the latter but are not alive. Nor are biopopulations and communities alive. Thus a forest teeming with plants, fungi, animals, and bacteria is not alive.

It follows that the class of biosystems is not included in that of biochemical systems, and that of biopopulations is not included in that of biosystems. What we do assert instead is that, just as biochemical systems compose biosystems, so the latter compose populations. Therefore we reject ontological reductionism, according to which biosystems are just a kind (subset) of biochemical systems, and biopopulations in turn a kind (subset) of biosystems. We assert instead that the components of members of the biopopulational level are members of the biosystem level, which are in turn composed of members of the biochemical level. This emergentist thesis is compatible with the program of explaining the higher in terms of the lower or, more precisely, the system in terms of its components (and the interactions among them). We shall come back to these matters in Sec. 4. (See also Bunge, 1977f.)

2. BIOFUNCTION

2.1 *Health and Death*

Systems, in particular organisms, can cope with their environment with varying degrees of success: recall from Definition 1.14 that the selection pressure exerted by the environment on a population of systems of some kind is a number lying between 0 (total success) and 1 (utter failure). But that concept of degree of adaptation was collective, nonspecific, and phenomenological: it does not carry over to an individual biosystem and it is not physiological. We need a separate concept of degree of adaptation of a given biosystem, not necessarily an organism. Here is one:

DEFINITION 3.5 Let *b* be a biosystem in an environment *e*. Then *b* is *healthy or normal in e* iff *b*, when in *e*, possesses all of the properties (can discharge all of the functions) listed in Postulate 3.1.

Remark 1 This is a physiological concept of health or normality, not a statistical one. It is possible for an entire population of healthy organisms to be wiped out by an environmental catastrophe, such as a severe drought. And it is also possible for most of the individuals of a biopopulation to be sick – a case of statistical normality and physiological abnormality. *Remark 2* Sickness is sometimes regarded as a form of adaptation. This is so in the case of infectious diseases, when the organism synthesizes antibodies that "fight" the invaders. But it is not so in other cases: there is nothing adaptive about arthritis, the obstruction of the bile canal, an infarct, stroke, or stress. Quite the contrary, these are glaring cases of failure to adapt. *Remark 3* Our definition is not limited to whole multicellular organisms: it may apply to cells or organs of a multicellular organism. Nevertheless in such cases it is preferable to restrict oneself to the specific functions discharged by such subsystems. More on this in Sec. 2.2.

The concept of health elucidated by the previous definition is qualitative. It is desirable to have also a quantitative concept of (degree of) health (or sickness) as well as of total health (or health in all respects). One obvious candidate is this: The degree of sickness of a biosystem in a given environment, in a given respect, at a given time, is its relative imbalance (or deviation from the physiological norm) in that respect at that time. Assuming optimistically that such normal values can always be found, we propose

DEFINITION 3.6 Let $F_i\colon B \times E \times T \to \mathbb{R}$ be the ith component of the state function of biosystems of kind B in environments of type E, and $\bar{F}_i$ the corresponding (physiological) normal value of F_i. Further, let $w = \{w_i \in [0,1] \mid 1 \leq i \leq N\}$ be a set of positive reals that add up to unity, and such that w_i measures the weight or relative importance (for the whole biosystem) of the property represented by the function F_i. (Cf. Vol. 3, ch. 2, Sec. 4.2, Definition 2.15 of property weight.) Then

(i) the *degree of sickness* of the biosystem $b \in B$ in the environment $e \in E$, in the ith respect and at time $t \in T$, is

$$s_i(b, e, t) = \left| F_i(b, e, t) - \bar{F}_i \right| / \left| F_i(b, e, t) + \bar{F}_i \right|;$$

(ii) the *degree of health* of the biosystem b in the environment e, in the ith respect and at time t, is the complement of $s_i(b, e, t)$ to unity, i.e.

$$h_i(b, e, t) = 1 - s_i(b, e, t);$$

(iii) the *total degree of sickness* and the *total degree of health* of organism b in e at t are

$$S(b, e, t) = \sum_{i=1}^{N} w_i s_i(b, e, t) \quad \text{and} \quad H(b, e, t) = 1 - S(b, e, t)$$

respectively.

The values of s_i and h_i are real numbers in the unit interval. And each is the same whether $F_i(t)$ equals $n\bar{F}_i$ or $(1/n)\bar{F}_i$, with n an arbitrary integer.

The conceptual interest of the above formula resides in its making sense only within a theoretical model of biosystems of the species of interest. Its methodological interest lies in its reminding us that scientific medicine is just a chapter of biology (including psychology in the case of man).

Finally we come to

DEFINITION 3.7 A thing is *dead* iff, having been a biosystem, it has ceased to perform all of the functions listed in Postulate 3.1.

Remark 1 We could not define death as maximal degree of sickness, or as minimal degree of health, because the concepts of sickness and health refer to living things (Definition 3.6). *Remark 2* Surely not all of the vital functions stop at the same time: death is a process rather than a point event. However, we could not define dying as the gradual cessation of biofunctions, because recovery is not unusual, whereas dying is irreversible. *Remark 3* Because all biosystems emerge from biochemical systems or even from whole cells rather than from scratch (i.e. from physical entities), the process of dying is not the inverse of the birth process. See Figure 3.3 *Remark 4* Life and death are not entities but states of, or processes in, biosystems. Death is the process of breakdown of a biosystem into an aggregate of chemicals. It starts when the biochemical reactions that maintain a steady state – namely the metabolic processes – come to a halt. In particular, no new enzymes necessary for the synthesis of biomolecules are produced and, in general, breakdown processes start to prevail over buildup processes. *Remark 5* That health and sickness, life and death are not entities but states of entities, is obvious from a systems theoretic viewpoint but not from alternative points of view. Indeed, sickness and death have been treated as entities, not only by primitive men but also by some medical and philosophical schools. Having reified and even personified death, it was possible to regard life either as a fight against Death or as the transit towards it. The anxieties elicited by such primitive views (contained in contemporary existentialism) are allayed by the naturalistic doctrine proposed by Epicurus, according to which dying is a process of dismantling of an organism.

2.2. *Function and Value*

Every cell, no matter how simple or primitive it may look at first sight, is composed of a number of biochemical systems of enormous complexity

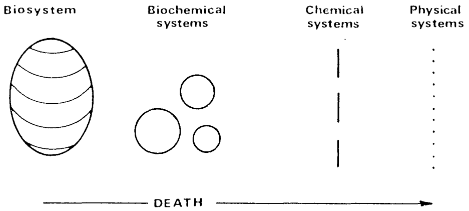


Fig. 3.3. The successive collapses of the state space of a dying biosystem: a gradual loss of properties and of systemicity.

which are linked to one another, and succeed one another, in various ways. If eucaryotic (nucleated), a cell consists of a nucleus and a cytoplasm. The former is composed of DNA and RNA threads, and the latter of subsystems such as ribosomes, mitochondria, chloroplasts (in plants), and cilia – none of which are alive.

Each of the nonliving subsystems of a cell performs certain activities, usually called 'functions', such as storing energy, synthesizing proteins, breaking down excess biomolecules, detecting imbalances, or transmitting signals. Such subsystems of the cell are therefore called 'functional units'. In multicellular organisms (metazoans) this division of labor often carries over to entire cells or even cell groups such as tissues and organs. In either case each subsystem of a biosystem, whether or not the former is alive, discharges certain functions, i.e. does certain things, that are peculiar to it, and others which are done also by subsystems of other kinds. For example, neurons not only generate and transmit signals but also synthesize ATP, nucleic acids, and proteins,

We elucidate these notions by means of the following definition, which employs the notion of a process elucidated in Ch. 1, Sec. 1.7:

DEFINITION 3.8 Let b be an organism, and $a \prec b$ a subsystem of b, of kind $[a]$. Further, call $\pi(a)$ the totality of processes or activities involving a during a certain period. Then

(i) any subset of $\pi(a)$ that includes any of the processes listed in Postulate 3.1 is a *biological function*;

(ii) the *specific biological functions* of a are those performed by a and its likes but not by any other subsystems of b. I.e.

$$\pi_s(a) = \pi(a) - \bigcup_{x \prec b} \pi(x), \quad \text{with} \quad a \neq x \prec b \ \& \ x \notin [a].$$

The biological functions of a subsystem of an organism are valuable or disvaluable to the organism as a whole. For example, the specific functions of the liver are to synthesize certain substances (e.g. albumin and cholesterol) and to store others (e.g. glycogen and iron). No mammal can live without a liver: the latter's value to the organism is maximal, though of course not greater than the values of the heart and the brain. On the other hand a mammal can live without a gall bladder although this organ markets (stores, concentrates, and delivers) the bile produced by the liver, and the bile is useful for digesting fats and other substances. That is, the value of the gall bladder, though positive, is far smaller than that of the liver. Finally, some subsystems of an organism perform disvaluable biological functions: think of tumors.

The above suggests the following measure of the biological value of a subsystem a of an organims b in an environment e: The value of a to b in e equals the degree of health of b in e when equipped with a, minus the degree of health of b in e when deprived of a. More precisely, with the help of Definition 3.6 we lay down

DEFINITION 3.9 Let $H_A\colon B \times E \times T \to \mathbb{R}$ be the health function for organisms of kind B, in environments of type E, when subsystems of kind A are acting in the B's (i.e. when $a \prec b$ and $\pi_s(a) \neq \varnothing$ for $a \in A$ and $b \in B$). Similarly, call $H_{\bar{A}}\colon B \times E \times T \to \mathbb{R}$ the corresponding health function when the A's are not acting in the B's (i.e $\neg(a \prec b)$ or $\pi_s(a) = \varnothing$). Then the *value of A for B in E* is the function

$V\colon A \times B \times E \times T \to \mathbb{R}$ such that, for any $a \in A$, $b \in B$, $e \in E$ and $t \in T$,

$$V(a, b, e, t) = H_A(b, e, t) - H_{\bar{A}}(b, e, t),$$

which is the *value of subsystem a for organism b in environment e at time t.*

Since health ranges between 0 and 1, biological values are comprised between -1 and 1. In particular

a is *maximally valuable* to b in e at t iff $H_A(b, e, t) = 1$ and $H_{\bar{A}}(b, e, t) = 0$;
a is *worthless* to b in e at t iff $H_A(b, e, t) = H_{\bar{A}}(b, e, t)$;
a is *maximally disvaluable* to b in e at t iff $H_A(b, e, t) = 0$ and $H_{\bar{A}}(b, e, t) = 1$.

Thus in humans the heart is maximally valuable, the normal appendix worthless, and a malignant tumor maximally disvaluable. And the values of other systems, such as the reproductive system, the thymus gland and the tonsils, vary with age.

The preceding definition allows one to elucidate another concept, this time of particular interest to surgeons and bioengineers:

DEFINITION 3.10 Two subsystems a and a' of a given biosystem b are *functionally equivalent* iff

(i) a and a' perform the same specific functions, i.e. $\pi_s(a) = \pi_s(a')$, and

(ii) a and a' have the same value during the time interval of interest, i.e. $V(a, b, e, t) = V(a', b, e, t')$ for t and t' in the given interval.

Remark 1 Some biofunctions, such as metabolism and cell division, are common to all organisms. Others are species-specific. For example photosynthesis is peculiar to plants, and the synthesis of enzymes that occurs in the digestion of meat is peculiar to carnivores. The grouping of organisms by biofunction clusters provides just as natural and fundamental a classification as evolutionary taxonomy; only, it is out of fashion. *Remark 2* The more numerous the functions an organism can perform, the better equipped it is to cope with environmental challenges. Multicellular organism that are highly differentiated will therefore be selected by rich and changing environments. Also, because they have spare cells, metazoans are covered against the accidental loss of some cells, so they have better chances of survival than organisms without such spares. (Recall the discussion of reliability in Ch. 1, Sec. 4.1.) These two factors help explain the prosperity of metazoans on our planet. In short, the terrestrial environment has been fit for complex and versatile organisms endowed with a high degree of redundancy. *Remark 3* Alternative definitions of biological value are of course thinkable. One alternative worth looking into is to equate biovalue with efficiency. Another is teleological: Valuable to an organism (or a species) X is whatever contributes to X's attaining its goals. We cannot adopt this alternative because we are not assigning goal-seeking activities to all organisms. Goals will be introduced in the next chapter, with reference to organism endowed with complex nervous systems. *Remark 4* As evolution proceeds, the value of some subsystems increases (preadaptation) and that of others decreases, sometimes to the point of vanishing altogether. In the latter case the subsystem may disappear (as in the loss of the ability to synthesize certain vitamins), or it may be kept functionless for hundreds of thousands of years. Thus it is possible, nay quite likely, that man has a number of fossil (functionless) proteins, genes, entire cells, and even cellular systems – i.e. subsystems that have long ceased to serve any biofunction. But this is very hard to prove. *Remark 5* It is just as likely that certain hereditary pathological conditions were valuable in the past or may yet become valuable. This is the case with sickle-hemoglobin, which is associated with resistance to malaria. Likewise certain congenital malformations, such as flat feet and poor sight, have spared the lives of many men in modern wars. (War selects the unfit.)

2.3. *Biocontrol*

All systems are under environmental control. And biochemical systems are also under internal – in particular enzymatic – control. (Recall Definition 2.17.) Biosystems are then controlled by their environment (in particular by the availability of precursors of biomolecules) and by their biochemical subsystems. In particular, enzymes and nucleic acids exert internal controls. Besides, multicellular organisms have additional control systems, such as the control of growth and size by specific hormones. In short, all organisms are subject to dual controls: internal and environmental, or self and external.

Naturally, the self-control of a biofunction is called *biocontrol.* Most biocontrols are of the feedback type. (See Appendix A, Sec. 1.3.) In such cases, if the output of a subsystem falls below (above) a certain normal value, the control system increases (decreases) the input in such a way that the imbalance is gradually redressed. (The normal output level is sometimes misleadingly called the set of *goal states* or *Sollwerte.* But there is no evidence of goal directed behavior except for certain features of the behavior of animals equipped with an advanced brain: see Ch. 4.)

When all the controls of an organism work properly, the organism attains or maintains a steady state or *homeostasis*: it stays in or near such a state regardless of external perturbations, provided these are bounded. However, homeostasis is usually a global result of frantic molecular and activity. Moreover there are a number of nonhomeostatic states in the life history of every organism, namely all those leading to and from homeostasis – such as the processes of birth and death, of sickness (or injury) and recovery from it. In these cases one speaks of *homeorhesis.*

Biocontrol starts at the molecular level. Thus DNA controls the synthesis of RNA, which in turn controls protein synthesis. (See Ch. 2, Sec. 2.1.) But DNA itself, far from being an unmoved prime mover, is under severe control. In fact, according to the Jacob-Monod hypothesis, genes are turned on or expressed only when such control is lifted. (It is likely that, at least in multicellular organisms, many genes are repressed or inactive most of the time and, like firemen, become activated only during emergencies.) In other words, only when the control is off the gene is free to participate in the synthesis of whatever protein(s) it may. And this is not the whole story: some enzymatic proteins in turn control RNA synthesis. (Furthermore, catalytic cycles formed by RNA molecules and enzymes may well exist: see Eigen (1971).) In short, genetic control is not ultimate and hierarchical. This is a point of philosophical interest in view of the fashionable claim that life is at the mercy of a bunch of selfish and smart genes.

In addition to the enzymatic and genetic controls, which regulate directly the thousands of biochemical reactions going on in the organism, there are highly specialized biocontrol systems, such as those that control acidity, temperature, growth, and motion. These are unicellular or multicellular systems, every one of which can be analyzed into a *sensor* and an *effector* (or *corrector*). The sensor evaluates both the input and the output, and "informs" the effector whether they are adequate. (Actually there is no need for the information concept at this level: all the sensor does is to trigger the effector as soon as it detects an imbalance.)

In a sense every concrete thing can "sense" or detect some other things. Thus an atom can detect light (i.e. react to it) as well as the proximity of other atoms. However, biocontrols are *specific*: some detect gravity (or rather deviations from the vertical); others light intensity; still others light frequency, or acidity differences, or temperature differences, and so on. Furtermore, whereas physical and chemical detectors usually detect absolute values or intensities, all biosensors detect *imbalances*, i.e. differences between actual and normal ("goal") values. This they achieve by being components of negative feedback loops. In short, biocontrols are specific and they correct imbalances.

Furthermore, unlike a bunch of physical detectors, the biocontrols in an organism constitute a network: they interact and, in the healthy organism, they do so synergically, i.e. harmoniously rather than independently from one another. (Such interaction may involve so little energy, and yet may have such important effects, that it is usually called 'informational'. See Appendix A, Sec. 1.4.) For example, the various internal secretion glands in a vertebrate are interconnected by hormones carried by the bloodstream, and the various parts of the nervous system keep in touch ("talk to each other") by synaptic contacts as well as by hormonal "messengers". In short, the control systems in an organism constitute signal (or informational) networks. Actually every organism, however primitive, is equipped with two interconnected signal networks: an internal network constituted by its internal biocontrols, and an interfacial network constituted by biocontrols that sense environmental changes. These two networks are coupled and between them they control all the functions of the organism.

This, then, is one of the peculiarities of organisms: namely that they have multiple control systems grouped into tightly integrated *signal networks*. Another peculiarity is that such networks function well – amazingly well in most cases. 'Well' does not mean either at full blast or else with maximal economy, but instead in whatever manner is most valuable to the organism as a whole. (Recall Definition 3.9 of biovalue.) For example, the rate at which any given protein is synthesized is neither minimal (maximal economy of resources) nor maximal (waste) but somewhere in between – above equilibrium and short of waste. This near optimality is peculiar to biosystems.

Unlike biology, physics and chemistry are dominated by extremal principles such as Hamilton's, i.e. basic law statements according to which, of all the conceivable processes of a given kind, only those are actualized that either minimize or maximize a certain magnitude, e.g. the action. (Cf. Appendix B, Sec. 1.4.) These principles "operate" of course for the physical components of a biosystem, but these components intertwine in such a manner that certain properties of value to the organism as a whole are near-optimized rather than either minimized or maximized. I.e. such near optima are not determined locally by each separate system but result from the interaction among the various subsystems of the organism. In fact since the various components of the control network of the whole organism are interdependent, the near optimization of certain local variables will hamper that of others. So, what would be best for each subsystem is unattainable. But the next best is attainable, and this is a compromise among a number of optima. This compromise may be called *overall* (or organismic) *near optimality* – not full optimality because no organism is perfect. (A perfect biospecies would not be subject to evolution.) Only the biological organization of *successful* species *tends* to optimality.

An organism may be said to possess an efficient biocontrol network if it is capable of redressing whatever imbalances may occur – i.e. if it can detect and correct them. If not, the organism becomes sick and may die. More precisely, every malfunction of a specific biocontrol system is a specific sickness. (The converse is false. For example infectious diseases, if overcome, are so by the proper functioning of the immune response system.) And if several biocontrol systems malfunction, or do not function synergically, a nonspecific (generic) sickness supervenes. Natural death is probably the final state in a process of generic (nonspecific) sickness.

Biocontrol has given rise to the belief that organisms are goal-seeking: that they act in everything so as to ensure their own preservation or at least that of their species. First the theory of evolution through natural selection, and in recent times the theory of control, should have corrected that illusion. In the first place self-control is a property of all biochemical systems – by virtue of their enzymatic activity – whether or not they are alive. And if system integrity were defined as equilibrium (steady state), then every biochemical system would be seen to behave *as if* it reached for its own preservation. Secondly, biocontrol is not always efficient: it falls short of optimality when the organism becomes sick and it fails utterly at death. Thirdly, biocontrol can fail, even in healthy organisms, when certain imbalances grow beyond bounds during catastrophes such as floods and droughts. (After all, the great majority of terrestrial biospecies have become extinct.) Fourthly, biocontrol is occasionally defective. For example, only a small part of the RNA synthesized in the cell nucleus migrates to the surrounding cytoplasm to control protein synthesis: most of it is broken down into nucleotides without ever being "used". Another example: many genes are never "expressed" or activated. Fifthly, genetic control is not so strict as to prevent mutations, most of which are deleterious. In any case, we need not invoke goal-directedness to explain why biocontrol is normally beneficial for the organism as a whole: the organisms not endowed with efficient control networks succumb before reproducing, and that's all.

2.4. *Development*

The history of any system can be represented as the trajectory of its representative point in a suitable state space for the system. (Cf. Ch. 1, Sec. 2.2.) In particular, the *life history* of an organism can be represented as a line in its state space. The part of the life history comprised between the beginning (birth) and the attainment of the adult state of the organism is called its *development*. (See Bonner (1974).) The development process of an organism has three main strands, each of which is taken care of by a bunch of components of the state function representing its possible states. These strands or aspects of development are:

(i) *growth*, or multiplication of cells of certain kinds (at uneven rates for different subsystems of the organism);

(ii) *morphogenesis*, or formation of new subsystems;

(iii) *differentiation*, or specialization of cells.

All three aspects of development proceed under a triple control: enzymatic, genetic, and environmental. The environment supplies or withholds the wherewithall, the enzymes make the transformation of environmental precursors possible, and the genes guide the entire process – provided their controls permit them to do so, as we saw in the previous subsection. (Some molecular biologists tend to play down the control effected by the environment. However, the latter can be dramatic, as illustrated by the sporulation of moulds that occurs in a poor environment.) One way of characterizing the salient features of the development process is by means of

DEFINITION 3.11 If *b* is a biosystem, then

(i) *b grows* iff (*a*) some DNA molecules in *b* replicate, and (*b*) the rates of RNA and protein syntheses in *b* are greater than the corresponding breakdown rates;

(ii) *b* is involved in *morphogenesis* iff some processes in *b* consist in the formation of new subcellular systems (*intracellular morphogenesis*), or new cells (*cellular morphogenesis*), or new organs (*organogenesis*);

(iii) *b differentiates* iff *b* is a multicellular organism such that (*a*) the number of specific functions of *b* increases as the number of cells increases, and (*b*) different specific functions are performed by different cells or multicellular systems.

Of these three aspects of development, morphogenesis has always seemed the most mysterious and the one most clearly indicative of design and purpose. To begin with, the very word 'morphogenesis' suggests a process whereby the formless (e.g. an egg) acquires a definite form (shape or spatial pattern). While this is so at the physical and chemical levels, as in the cases of the generation of waves in a homogeneous fluid, or the growth of crystals in solutions, it is not so at the biolevel: here morphogenetic processes are those where new forms emerge out of old. (For examples and models of self-organization processes at various levels, see Glansdorff and Prigogine (1971), Gierer and Meinhardt (1972), Glass (1973), Tyson (1976), and Nicolis and Prigogine (1977).) Particularly in embryology every developmental phase is structured rather than amorphous. Moreover, shape change is not the primary or driving feature of development but an outcome of a profound process such as the diffusion of the products of an enzymatic reaction, or the multiplication of cells under the control of morphogens – enzymes, hormones, and genes. (Classical morphology was restricted to describing the net outcome of such processes: it neither attempted nor could explain them by hypothesizing biochemical and cellular mechanisms. Likewise the pretentious work of

Thom (1972), with its total disregard for morphogens, physical constraints, natural selection, chance, and history, does not go beyond classical morphology except for the addition of a mathematics of dubious utility.)

Morphogenesis may be regarded as a particular case of self-assembly of molecules, organelles, or cells, as the case may be. Like other self-assembly processes, the morphogenetic ones are guided by both internal forces and environmental resources and circumstances. The formation of an organelle, a cell, or an organ is controlled by genes and other morphogens in conjunction with the *milieu intérieur* and the environment. The DNA and RNA molecules, the hormones and other morphogens control the chemical reactions and the diffusion proceses, as well as the assembly and dismantling of the various units, but nothing gets done unless the environment supplies the materials and the energy.

A self-organization process can start spontaneously in an initially homogeneous system of reactants and morphogens. Thus a local chance fluctuation in their relative concentration can diffuse throughout the medium, ending up in a differentiated system (Turing, 1952). However, reaction-diffusion processes are not the only morphogenetic mechanisms, nor are they typically biological. Alternative mechanisms are: (*a*) sorting processes (grouping of functionally similar units, as in the case of polyribosomes), (*b*) cooperative interactions among molecules or cells of different kinds, (*c*) packing, folding, and other mechanical effects of accretion and external pressure; (*d*) enzymatic syntheses of morphogens proceeding until a new entity is formed – and inhibited thereafter.

Let us take a quick look at processes of the first kind, i.e. sorting processes (Steinberg, 1963; Mostow (Ed.), 1975). Let C be a set of cells of two kinds, A and B. Assume that all these cells are motile and that like cells adhere to one another more strongly than do unlike cells. Moreover, suppose that the cells of type B adhere to one another less strongly than those of type A. In sum, assume that the adhesion works (or energies) compare like this:

$$W_{\mathrm{A}} > W_{\mathrm{B}} \geq W_{\mathrm{AB}},$$

where W_{A} measures the adhesion of cells of kind A, W_{B} that of cells B, and W_{AB} that of unlike cells. One can imagine, and indeed prove rigorously, that the cells clump together and sort out all by themselves. In particular, if $W_{\mathrm{AB}} = 0$, the sorting will be complete: two separate tissues will ensue. See Figure 3.4.

Although the details of such self-assembly processes are far from being

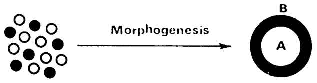


Fig. 3.4. The formation of an onion-shaped organ constituted by two concentric spheres composed of cells of different kinds, A and B, according to Steinberg (1963).

known, there is no excuse for regarding them as unknowable. Morphogenetic processes look mysterious only if one adopts either a vitalistic view or a primitive version of mechanism – one according to which matter is a mere aggregate of inert particles. But biology has outgrown both extremes: it has no use for supramaterial entities or forces, and it has learned that molecules, particularly carbon compounds, can stick together in numerous ways, forming systems of great complexity and often also of great beauty.

One outcome of morphogenesis in multicellular organisms is differentiation. The degree of differentiation of a system may be taken to equal the total number of specific functions performed by its various subystems. Now, for a complex system to keep together, its various subsystems must be well coordinated. In other words, while systemicity in general requires the existence of bonds among the system components, *organic systemicity* requires that these bonds be well coordinated, so that nearly every system component be valuable to the system as a whole. In particular, not only the right reactants must be present under favorable circumstances to their reactions, but the latter must proceed at the proper pace and in the proper sequence; and no component must block the operation of the other components – unless such blocking be of value to the organism as a whole. All this is compressed into the well worn platitude that "The greater the differentiation of an organism, the greater its coordination and integration". (Needless to say, it is not the case that coordination is a goal of development, but rather that the organisms that fail to achieve efficient coordination perish.)

2.5. *Reproduction and Heredity*

One of the properties of organisms is that they are capable of reproducing. (Sexually reproducing organisms do not reproduce unless they mate, but most of their cells divide. And neurons do not normally divide but other cells, linked to them, do.) Moreover organisms reproduce "true to type" i.e. their offspring belong to the same species. (Speciation takes many generations because it is a result of cumulative mutations and successive selections.)

The root of reproduction is the chemical process of DNA replication. (See Ch. 2, Sec. 2.1.) This explains the invariant aspect of heredity, particularly in asexually reproducing organisms. Sexually reproducing organisms are far more variable because of the recombination (nearly random shuffling) of parental genes. In either case the genetic system, or genome – which is composed of DNA molecules – is "in charge" of heredity. Organisms are not pulled by goals but pushed by the genetic memory of the past – pardon the metaphor.

The genetic system, at one time thought to be immutable, is known to be subject to mutation. The mutation mechanism is molecular. Consider the DNA molecules in a cell. As we know, these molecules can replicate. Now, during the replication process accidents ("mistakes") of two kinds can occur: (*a*) the daughter molecular strings may differ from the parent string in the order or sequence of the nucleotides (i.e. the daughter DNA molecules are isomers of their parents), or (*b*) one or more nucleotides may be either deleted or inserted (substitution). And, because DNA controls protein synthesis, such "mistakes" are likely to ensue in the synthesis of proteins of kinds different from those contained in the parental cells(s) – i.e. the mutant DNA can "code for" (control the synthesis of) new proteins. And the latter make it possible for the offspring to interact in new – not necessarily more efficient – ways with its environment. The end result is a mutant organism – one that may or may not be viable. See Figure 3.5.

Mutations are not frequent but they are bound to occur either in large populations or along many generations. And the number of possible mutations, hence of possible new proteins, and therefore of possible organisms more or less different from their ancestors, is staggering. (Speak of the poor "possible worlds" imagined by philosophical logicians!) Mutation is indeed an inexhaustible source of biotic novelty. And because mutation occurs by chance and irrespective of its value or disvalue to the organisms, chance has been said to be the sole source of newness in the biosphere (Monod, 1970, p. 127). But this is an exaggeration. The chemical affinities at the root of the nucleotide sequence are hardly chancy, and the selective action of the environment has causal components as well as random ones. (See Sec. 3.)

The large number of possible mutations explains the proteinic uniqueness of each organism. Equivalently: it accounts for many individual differences within any given species – or what is misleadingly called "indi-

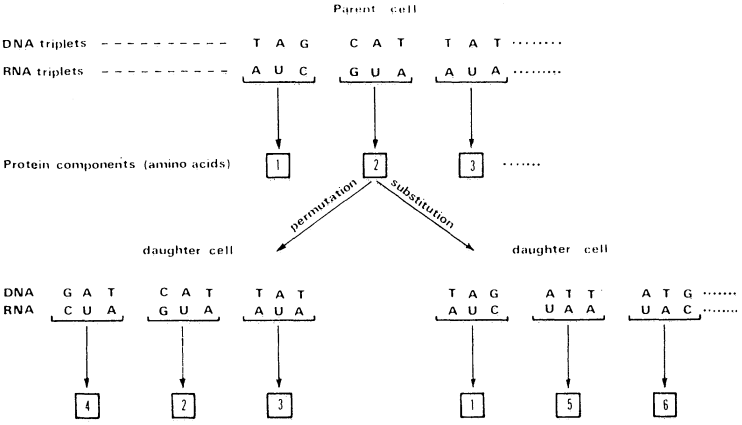

Parent cell
DNA triplets
T A G
C A T
T A T
RNA triplets
A U C
G U A
A U A
Protein components (amino acids)
1
2
3
permutation
substitution
daughter cell
daughter cell
DNA
RNA
G A T
C U A
C A T
G U A
T A T
A U A
T A G
A U C
A T T
U A A
A T G
U A C
4
2
3
1
5
6

Fig. 3.5. Two kinds of mutation: by permutation and by deletion (or insertion) of nucleotides.

vidual variability". It also explains the large number of biospecies. And the limited environmental resources, with the accompanying competition for them, explains why biopopulations do not grow exponentially (except for limited periods). Finally, changes in environmental conditions – particularly important changes such as glaciations and continental drifts – explain why (*a*) most biospecies that have been on Earth have become extinct, and (*b*) the number of species can jump up or down during certain brief periods. (One such revolutionary period was the beginning of the Cambrian, when many species of multicellular organisms emerged.)

In view of the fashionable claim that whatever an organism is and does is programmed in its genome, it may be helpful to insist that what is inherited is a bit of genetic material carrying certain potentialities. These potentialities may or may not be actualized in the course of the organisms' life, according to the possibilities offered by the environment. (For the concepts of potentiality and actuality see Vol. 3, Ch. 4.) We may as well be more precise and adopt

DEFINITION 3.12 If b is an organism, then

(i) the *genome* of b = the potentiality of b at birth;

(ii) the *phenome* of b = the set of molar (global) properties of b in adult state.

While geneticists focus on the genome, others focus on the phenome. Both visions are myopic, for only a suitable genetic equipment allows the organism to cope with its environment, and only a suitable environment enables the genetic traits to "express" themselves (i.e. to become activated). Thus, while every organism inherits its genome, it evolves its own phenome in a given environment; the phenome depends then upon both the genome and the environment. Given the variability of both, evolution should come as no surprise. But evolution will concern us in Sec. 3, after we look at the philosophical background of the contemporary doctrines on development and heredity.

2.6. *Development, Heredity, and the Metaphysical Tradition*

Contemporary studies of development and heredity have been influenced, mostly unwittingly, by ancient metaphysical ideas on the relations between substance and property, on the one hand, and between potency and actuality on the other. In turn, the results of such studies should be allowed to influence our general ideas on the substance-property and the potency-actuality pairs.

Consider first the organ-function relation as a particular case of the substance-property relation. Every biofunction is the function of some organelle or some organ, but the converse is false: i.e. there may be organs that have no biofunction (the vestigial or "fossil" organs), and others that perform specific functions only at certain stages of development. In the latter case the organ sometimes precedes the functioning organ – or, as the misleading idiom has it, "organ precedes function". For example the eye, the reproductive system, and large parts of the brain of a cat are formed before they function fully.

The temporal precedence of the nonfunctional organ over the fully functional one would seem to refute the Platonic doctrine that form (or essence or idea) precedes existence. Not so, for the Platonist could rejoin that the form (function) is inserted into the organ when needed. What does refute the Platonic doctrine of forms is the well confirmed generalization that there is no form (function) without organ and, in general, no property without substance. (Cf. Vol. 3, Ch.2, Sec. 5.2.) The controversy between the substantialists ("organ determines function") and the functionalists ("function determines organ") is a particular case of the age-old metaphysical controversy over substance and property. This controversy dissolves upon realizing that neither determines the other. All that happens in the case of organs that do not perform their specific function during the entire life of an organism, is that their nonfunctional states precede in time the functional ones.

A second point where developmental biology and the theory of heredity make contact with metaphysics is over the question of preformation *vs.* emergence. There have been two classes of solution to the problem of explaining the development of an organism from egg through adulthood to final breakdown. One is to assume that all of the properties and patterns that appear in the course of development are just the unpacking or unfolding (*development*) of what was contained in potency (*preformed*) in the egg. Or, to put it into modern parlance, the assumption is that the genome contains a full set of "instructions" steering the organism through the successive stages of its development. This is the *preformationist* hypothesis held by Aristotle and Leibniz, and resurrected by some contemporary scientists who believe that biology is fully reducible to molecular biology. Preformationism denies real novelty, in particular the emergence of new properties that accompanies self-organization (Ch. 1, Sec. 3.2); and it minimizes or even ignores the role of the environment.

The partisans of *epigenesis* hold, on the contrary, that development is an epigenetic or creative process in which novelty occurs at every stage. They maintain that the organism starts at a low state of organization and acquires new components and properties, as well as loses others as it develops. Epigeneticism is of course the hypothesis underlying all of classical embryology as well as the contemporary studies on the origin of life.

The molecular biology revolution has all but eclipsed epigeneticism. What is fashionable nowadays is genetic preformationism. Yet fashion need not concide with truth. The claim that development is just the deployment or implementation of a blueprint contained in the genome is no more illuminating or correct than holding that it is guided by an immaterial *élan vital* (Bergson, 1907) – or that it consists ultimately in the embodiment of Platonic self-existing geometric shapes (Thom, 1972). Surely development can be neither purely preformative nor purely epigenetic, if only because the environment supplies the raw materials for it and weeds out the maladapted organisms, as well as offers chances to the preadapted ones.

It seems that each view, either preformationism or epigeneticism, has alternatively accelerated and blocked the development of biology. Preformationism may have stimulated the purely descriptive phase of embryology and the beginnings of molecular biology, whereas epigeneticism may have stimulated experimental embryology (particularly investigations into the mechanisms of morphogenesis) and ecology. But neither view explains anything because each is just a programmatic hypothesis or a framework for discussing development and heredity. Indeed, far from being full-fledged theories, preformationism and epigeneticism are ontological hypotheses summarizing entire bodies of ideas, where sound data and conjectures are inextricably mixed up with metaphysical fossils.

It seems quite possible that the next breakthrough in the fields of development and heredity will come from a sort of merger of the two frameworks, one harmonizing two lines of research: (*a*) the search for specific development mechanisms (e.g. the mechanical constraints each cell is subjected to, the synthesis and action of morphogenes, and the interactions with the nutrient medium), and (*b*) the genetic "programming" of development possibilities – potentialities that are "encoded" in the genome in the sense that the latter contains, not "instructions", but some of the key reactants and templates required for the occurrence of certain biochemical reactions. But let us not attempt to write the history of the future.

3. EVOLUTION

3.1. *Adaptation*

A salient feature of our biosphere is its rich variety. There are at least three million biospecies (Dobzhansky, 1970), and no two strictly identical individuals in any of them. (Even so-called identical twins are somewhat different.) This variety is a result of four billion years of evolution. And the evolutionary process is in turn the net effect of a combination of genic changes (mutations and recombinations) with environmental selection. The process seems to be continuing: there is no reason to believe that it stopped the moment man became aware of it. In fact the extinction of a number of species has been certified during our own lifetime, and it is likely that new species are forming right now. In short, bioevolution is revolutionary – or creative (Bergson, 1907), or emergent (Lloyd Morgan, 1933).

Anthropomorphically speaking, evolution is a trial and error, or success and failure, story. Success is called *adaptation*, and failure *maladaptation*. A biopopulation of well adapted individuals thrives or expands, whereas one of maladapted individuals declines. This looks obvious but ceases to be so as soon as one stops and examines the very term 'adaptation': he then realizes how tricky it is (Williams, 1966). In fact, the word 'adaptation' is ambiguous, as it designates at least three different concepts, namely the following:

A_1 = *suitability* of a subsystem (organelle, organ, etc.) to a function, or high biovalue of the subsystem for the entire organism;

A_2 = *adjustment* of the organism to its environment;

A_3 = *fertility* of a biopopulation.

That these are indeed different concepts, is realized upon displaying their structure. A_1 is a relation between subsystems and organisms, A_2 a relation between organisms and environments, and A_3 an intrinsic property of organisms. A methodological examination confirms the difference. In fact the way to find out whether something has the property A_1 is to conduct certain biochemical or physiological experiments or observations in a constant environment. By varying environmental factors we can ascertain whether an organism possesses property A_2. And the criterion for A_3 is the count of offspring.

In sum, we must be reconciled with the fact that 'adaptation' is an ambiguous term. This seems obvious, yet it is overlooked by those who claim that only population genetics uses "the correct" concept, namely A_3. And it is equally overlooked by those who claim that the Darwin–Spencer principle of the *survival of the fittest* is tautological. Indeed, survivorship or fertility (A_3) is not the same as fitness or adjustment (A_2) but an indicator of the latter. The logical relations among the three concepts are these: $A_3 \Rightarrow A_2 \Rightarrow A_1$, i.e. subsystem biovalue is necessary for adjustment, which is in turn necessary for fertility.

It goes without saying that adaptation of any kind is time-dependent. I.e., A_1, A_2 and A_3 vary in time, either because of internal changes, or environmental changes, or both. The case of sudden environmental changes is particularly interesting, for they are bound to favor some biopopulations while hurting others. In the first case one speaks of *preadaptation*, in the second of *premaladaptation*. In keeping with the above distinction among three adaptation concepts, we distinguish three preadaptation types:

PA_1 switch of function(s) to other(s) of greater biovalue;

PA_2 improved adjustment to the environment;

PA_3 greater fertility.

As with the case of adaptation, the first is necessary for the second, which is in turn necessary for the third, i.e. $PA_3 \Rightarrow PA_2 \Rightarrow PA_1$.

Neither of the three kinds of adaptation or preadaptation calls for teleological considerations. The adapted or preadapted organisms may survive and reproduce, the maladapted ones don't, and that is the end of the story. To invoke adaptive *plans* at all evolutionary stages (Holland, 1975) is to indulge in anthropomorphism. There is not a shred of evidence for such master plans except in the case of man, and no serious biological theory involves them. Instead of imagining adaptive plans we should hypothesize and check adaptation mechanisms.

3.2. *Adaptation Mechanism*

Random mutation and recombination give rise to a great variety of unequally adapted organisms. The environment selects the best adapted amongst them and thus steers evolutionary change. Hence the environment is just as creative as the genetic material. But, whereas the latter varies randomly, the environment exerts a directional action. (Directional, not goal-directed – i.e. changing in a given direction but not directed towards a goal, let alone by a goal.) That, in a nutshell, is what the modern theory of evolution is all about.

A classical experiment confirming this theory is the following (Waddington, 1959). Larvae of fruit flies (*Drosophila*) are exposed to diverse environmental changes, such as excess salt, or heat, or ether vapour. If fed salted food, many of the flies die. The survivors happen to exhibit an enlargement of certain papillae conjectured to be involved in the regulation of salt concentration in the body. (The hypothesis is that this is a preadaptation, i.e. that those individuals that happen to be born with the more efficient salt controls, i.e. which are preadapted to the new circumstances, survive.) The 20th generation is definitely more resistant to salted food than the unselected or control group: roughly twice as many selected flies survive when given food with 5 percent salt added. The new character, i.e. the ability to cope with abnormally high salt concentrations, does not appear at one stroke but is instead the result of a gradual accumulation brought about by genic changes in conjunction with environmental pressures.

No master plan guides the gradual adaptation of organisms to their environment – until the latter changes unfavorably and maladaptation sets in, and eventually the population is decimated or even wiped out. Indeed, adaptation can be explained as the outcome of a random process powered by genic variation (mutation or recombination) and guided by the environment. To clarify this point consider a single property P, such as the size of the salt control system of *Drosophila* in Waddington's experiment. (Surely such a restriction to a single property is artificial, since every property is lawfully related to some other property of the same system. However, paying attention to a selected property does not entail denying the existence of other properties.) Make the following assumptions about P:

(i) P has a symmetric (e.g. bell-shaped) distribution in each population of the given lineage;

(ii) the successive generations do not overlap – so that the evolution of the population, as far as P is concerned, can be represented as a denumerable sequence of symmetric distributions;

(iii) there is an optimal value P_f of the given property, i.e. the individuals that come closest to possessing this value of P have the greatest survival chance. For the sake of definiteness assume that P_f lies somewhere to the right of the value P_i of the center of the initial distribution: see Figure. 3.6.

Variation and selection ensure that a gradual shift of the distribution in P to the right will occur as long as the optimal value P_f has not been attained. Indeed only those organisms with a P value close to the optimal P_f will be fit and they will tend to pass that characteristic on to their progeny – not because they want to secure the preservation of their species but because of the blind molecular process at the root of reproduction. As adaptation increases, survivorship increases and spread – i.e. individual variability – decreases. Where the vitalist sees a master plan or "teleonomic project", the evolutionist sees only a random process guided by natural selection. (Cf. Levins, 1968, Ch. 2; Dobzhansky, 1974.)

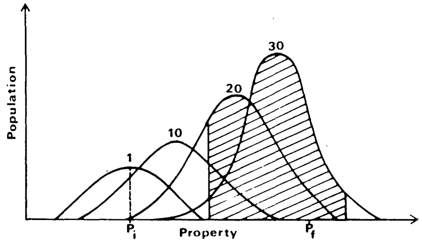


Fig. 3.6. Distribution of P-ness along 30 generations of organisms of a given lineage in a constant environment. The shaded areas represent fractions of survivols. In the beginning a tiny fraction of the total are adapted. At the end survivorship has increased enormously but variability has decreased. Imaginary and exaggerated.

3.3. *Evolution and Coevolution: basic ideas*

We have been referring to evolution without defining the concept. We do not have to look very far: the concept of bioevolution is just a specification of the general concept of history of a system (Ch. 1, Sec. 2.2) to biopopulations (Definition 3.4). That is,

DEFINITION 3.13 The *evolution* of a biopopulation is the history of the latter. I.e. if x is a biopopulation, then

$$\text{The evolution of } x = h(x).$$

What evolve, then, are biopopulations regarded as concrete systems spread out in space and time, rather than either individuals or species. (Individuals develop and species can succeed one another.) If we focus on qualitative changes and disregard intermediate steps, we can construe an evolutionary line as a succession of species or lineage (Definition 1.17).

Before Darwin proposed his theory of evolution by natural selection, bioevolution either was denied or was conjectured to have proceeded in a different way – e.g. by individual adaptation and the inheritance of acquired characters (Lamarck). We adopt of course the core of the Darwinian theory:

POSTULATE 3.4 All populations evolve, and the evolution of each is a process whereby spontaneous (genic) changes ensue in unequally adapted individuals, amongst which the environment selects the best adapted ones.

Selection is called *natural* if it is spontaneous, and *artificial* if controlled by man.

Because biopopulations are not isolated from each other but interact with other biopopulations forming ecosystems, the evolution of each must affect that of others. This process has a name of its own:

DEFINITION 3.14 Two or more populations *coevolve* iff the evolution of each affects that of the other I.e. if x and y are biopopulations, then x and y coevolve iff $h(x \mid y) \neq h(x)$ and $h(y \mid x) \neq h(y)$.

But we just said that every change occurring in any single biopopulation is bound to affect the biopopulations with which it interacts – in particular cooperates or competes. We obtain thus, albeit informally, a result important enough to be dignified with the name

THEOREM 3.1 Every biopopulation coevolves with every other biopopulation in the same ecosystem.

This is why it has been held that the unit of evolution (hence the referent of the theory of evolution) is not so much the monospecific population as the entire ecosystem (Dunbar, 1960). And why not the entire biosphere? These heresies are common sense in our systemic ontology, for any change in one system component is bound to affect every component to which it is coupled. (The only way to escape Theorem 3.1 is to deny that there are ecosystems and so to fly in the face of evidence.) The moral for the building of mathematical models of evolution is clear: Start by modeling the evolution of one biopopulation; proceed to model the coevolution of the (sub)system composed of the given biopopulation and its closest partner(s) or rival(s) – e.g. rock fish and the algae they eat; and aim at modeling the whole ecosystem of interest. Easier said than done!

In view of Definition 3.13, every evolutionary process can be represented as the trajectory of the representative point of the state of a biopopulation (or ecosystem) in its state space. However, it is not at all obvious how to choose the components of the state function spanning that state space. (Cf. Lewontin, 1974.) One possibility is to choose the gene frequencies (actually probabilities) as the coordinates or components of the state function. This is the course adopted by population genetics. It has not been too fruitful because selection operates on phenomes not genomes. (What the environment tolerates or wipes out are whole organisms, not just genomes.) The other extreme is of course to choose only molar or phenotypic variables as components of the state function. But this would leave the primary source of individual variability in the dark. And neither of the above strategies makes room for environmental variables. So we are led to a third approach, namely taking the census counts of the various subpopulations of a given biopopulation in the course of time, and viewing them as the net effects of genic variability and environmental selection. This will be done in the next subsection – with apologies to the biological profession for encroaching on their territory.

3.4. *A Schema of the Evolutionary Process*

We propose to sketch a simple model of evolutionary processes (Bunge 1978a). Consider a biopopulation of organisms such as bacteria of some kind, that reproduce by binary division. Assume that each individual lives one time unit, at the end of which it divides simultaneously with all the other members of the population. Assume also that mutants appear at each generation and that they can be grouped into distinct varieties. Call V_i the ith variety, where $1 \leq i \leq n$, and count the population of each variety at each generation. Further, call N_t^i the population of variety V_i at time (generation) t. If the fraction of each generation that dies without leaving offspring is $d_i < 1$, then the population of variety V_i at time $t + 1$ is

$$N_{t+1}^i = 2(1 - d_i)\, m_{ii}\, N_t^i, \tag{3.1}$$

where m_{ii} is the probability that division be not accompanied by mutation, and the factor 2 accounts for the division of each organism into two new individuals.

The previous formula assumes that all of the offspring of generation t belong to the same variety. In point of fact the number of individuals of variety V_i at time $t + 1$ will differ from the above by two terms: the fraction leaving the given variety and the fraction entering it, in both cases by mutation. Call m_{ij} the probability per unit time of a mutation from variety V_i to variety V_j. These probabilities are subject to the normalization condition

$$\sum_j m_{ij} = 1 \quad \text{for every } 1 \leq i \leq n. \tag{3.2}$$

The average number of defectors from variety V_i between times t and $t + 1$ will be

$$D^i = 2 \sum_{j \neq i} m_{ij}(1 - d_i)N_t^i. \tag{3.3}$$

And the converts to variety V_i in the same time interval number

$$C^i = 2 \sum_{j \neq i} m_{ji}(1 - d_j)N_t^j, \tag{3.4}$$

for they will be drawn from all varieties but V_i. The factor 2 in D^i and C^i shows that we are making the additional hypothesis that mutants come in pairs – i.e. that if an organism is mutant so is its sister.

Collecting (3.1), (3.3) and (3.4) we obtain the net average population of variety V_i at time $t - 1$:

$$\begin{aligned} N_{t+1}^i &= 2m_{ii}(1 - d_i)N_t^i - D^i + C^i \\ &= 2\left[m_{ii}(1 - d_i) - \sum_{j \neq i} m_{ij}(1 - d_i)\right]N_t^i \\ &\quad + 2 \sum_{j \neq i} m_{ji}(1 - d_j)N_t^j. \end{aligned} \tag{3.5}$$

Because of the normalization condition (3.2) of the mutation probabilities, the total probability of mutation from V_i is

$$\sum_{j\neq i} m_{ij} = \sum_{j} m_{ij} - m_{ii} = 1 - m_{ii}. \tag{3.6}$$

Introducing this value into (3.5) we are finally left with

$$N^i_{t+1} = 2(2m_{ii} - 1)(1 - d_i)N^i_t + 2\sum_{j\neq 1} m_{ji}(1 - d_j)N^j_t. \tag{3.7}$$

This system of n linear difference equations summarizes our model of evolution. The solution to these equations can be obtained as follows. Set

$$\begin{aligned} a &= \|a_{ij}\| = \|2(2m_{ii} - 1)(1 - d_i)\delta_{ij}\| \\ b &= \|b_{ij}\| = \|2m_{ij}(1 - d_i)(1 - \delta_{ij})\|, \end{aligned} \qquad N_t = \begin{Vmatrix} N^1_t \\ N^2_t \\ \vdots \\ N^n_t \end{Vmatrix}, \tag{3.8}$$

where $\delta_{ij} = 1$ iff $i = j$ and 0 otherwise. (The transpose of N_t is the value of the state function of the system at t.) The system (3.8) can be condensed into the matrix difference equation

$$N_{t+1} = (a + b^T)N_t, \quad \text{where} \quad \|b_{ij}\|^T = \|b_{ji}\|. \tag{3.9}$$

Iterating (3.9) we find the general equation of evolution

$$N_{t+k} = (a + b^T)^k N_t, \quad \text{where} \quad k \in \mathbb{N}. \tag{3.10}$$

The solution to this equation is

$$N_t = (a + b^T)^t N_0, \tag{3.11}$$

where N_0 is the column matrix of the initial populations of the n varieties being considered.

Suppose that all of the entries of the evolution matrix $a + b^T$ are positive, and that initially a single variety is present. Then all the other varieties can emerge simultaneously at the next generation, and the mutation-and-selection process proceeds relentlessly. Every one of the n varieties prospers once emerged, and the prosperity of each is conducive to that of all. (The only qualitative change in this process can be induced by a sudden environmental change causing some death rates to jump to unity.) But in general the process will not be explosive: while some entries of the evolution matrix may be positive, others may be zero or negative. An overall negative matrix corresponds of course to extinction.

As long as the death rates d_i and the mutation probabilities m_{ij} are not specified, the model is a phenomenological skeleton. Let us proceed to flesh it out.

3.5. *The Underlying Mechanism*

Let us now specify the architecture of our evolving organisms as well as their mutability (the m_{ij}) and viability (the d_i). We assume that the n varieties of our biopopulation differ by the composition or arrangement of a fixed number of modules or basic units that can be repeated. For the sake of simplicity we suppose that there are just two kinds of module, which we call 0 and 1. Thus if each entity is composed of four such modules arranged differently, we obtain a total of $n = 2^4 = 16$ varieties, namely 0000, 1000, 0100, 0010, 0001, 1100, 0110, 0011, 1010, 0101, 1001, 1110, 0111, 1011, 1101, and 1111. The general formula for a variety whose members are composed of m modules is $abc\ldots m$, where each letter is a binary digit. So much for architecture.

Every change in at least one of the m binary digits occurring in the formula for a variety is a mutation. Thus a mutant will differ from its parent organism either by the total number of 0's (or 1's) or by the order of 0's or 1's. Thus in the above example the transitions 0000 → 1000, 0100 → 0110, and 0111 → 1111 are mutations and, at the same time, changes in variety. Of course not all mutations are equally probable. We shall determine certain relations among mutation probabilities with the following considerations on viability.

Enter the environment. We lump all of the environmental factors into a single real valued function e with values in the $[-1, 1]$ real interval. And we assume that the probability that an organism belonging to variety V_i survives up to and including division is symmetrical about the origin. More precisely, we assume that the death rates d_i in (3.1), far from being constant, depend upon the environmental variable e in the following way:

$$1 - d_i = \begin{cases} v_i(1 - e^2) & \text{for } -1 \leq e \leq 1 \\ 1 \text{ otherwise}, \end{cases} \tag{3.12}$$

where $0 \leq v_i \leq 1$ for each variety V_i.

Since v_i is characteristic of each variety and is environment-independent, we call it the *adaptedness* of organisms of the variety V_i. The factor $(1 - e^2)$ represents the selection pressure effected by the environment. A substantial increase or decrease in e will have a considerable impact on the fraction $1 - d_i$ of a given population surviving until the next division. In particular, if a catastrophe occurs ($e = \pm 1$), no survivors are left, i.e. $d_i = 1$ for all i. On the other hand if the environmental conditions are optimal ($e = 0$), then the fraction v_i of individuals of variety V_i survive long enough to divide. The variable e is, in sum, the causal or nonstochastic component of evolution.

We introduce now an assumption that will allow us at least to rank the viability values. Our hypothesis is that the more varied the composition of an organism the better its chances of survival. Thus in the case $m = 2$, where we have only the varieties 00, 01, 10, and 11, we set

$$\begin{aligned} v_1 &= v_4 \equiv v && \text{for 00 and 11} \\ v_2 &= v_3 = \alpha v && \text{with } \alpha > 1 \text{ for 01 and 10.} \end{aligned} \tag{3.13}$$

Finally we relate viability to mutability, namely thus: we postulate that a mutation is the more probable the less viable its product is. (This assumption is in keeping with the well known generalization that most mutants are lethal.) More precisely, we assume

$$m_{ij} \geq m_{ik} \quad \text{if and only if } v_j \leq v_k. \tag{3.14}$$

We can now build the matrix of the mutation probabilities in the simple case where $m = 2$, whence $n = 4$. In this case the viabilities are those indicated in (3.14). Hence the reproduction chances are

$$\begin{aligned} 1 - d_1 &= 1 - d_4 = v(1 - e^2) \\ 1 - d_2 &= 1 - d_3 = \alpha v(1 - e^2). \end{aligned} \tag{3.15}$$

The hypothesis (3.14), that most mutations are deleterious, together with the assumption (3.13) on the adaptedness values, entails

$$m_{41} = m_{14}, \qquad m_{32} = m_{23} = m \tag{3.16}$$

(where this m is not to be confused with the number of kinds of modules per organism, which we have taken to be 2), and

$$\begin{aligned} m_{21} &= \beta_1 m_{12}, & m_{31} &= \beta_1 m_{13} \\ m_{24} &= \beta_2 m_{42}, & m_{34} &= \beta_2 m_{43}, \end{aligned} \tag{3.17}$$

with $\beta_1, \beta_2 > 1$. We simplify by setting $\beta_1 = \beta_2 = \beta$, and so are left with

$$\begin{aligned} m_{21} &= \beta m_{12}, & m_{31} &= \beta m_{13} \\ m_{24} &= \beta m_{42}, & m_{34} &= \beta m_{43}. \end{aligned} \tag{3.18}$$

Since module permutations are far less radical than changes in composition, we make the further simplifying assumptions that

$$m_{12} = m_{13}, \qquad m_{24} = m_{34}. \tag{3.19}$$

Similary, we posit that the variety 01 is just as stable as 10, and that 00 and 11 are equally stable:

$$m_{22} = m_{33}, \qquad m_{11} = m_{44}. \tag{3.20}$$

Our final assumption is that the mutations between the extreme varieties 00 and 11 proceed via the intermediate species 01 and 10, i.e.

$$\begin{aligned} m_{14} &= m_{12}\cdot m_{24} + m_{13}\cdot m_{34} \\ m_{41} &= m_{42}\cdot m_{21} + m_{43}\cdot m_{31}. \end{aligned} \tag{3.21}$$

A first consequence of the above assumptions is

$$\begin{aligned} m_{21} &= m_{31}, \qquad m_{42} = m_{43} \\ m_{14} &= 2m_{12}\cdot m_{24} = m_{41} = 2m_{21}\cdot m_{42}. \end{aligned} \tag{3.22}$$

Finally we use the normalization condition (3.6) which, together with the previous equalities, yields the mutability matrix in terms of m_{11}, m_{22} and m:

$$\|m_{ij}\| = \left\| \begin{matrix} m_{11} & \dfrac{(1-m_{11})}{(3-m_{22}-m)} & \dfrac{(1-m_{11})}{(3-m_{22}-m)} & (1-m_{11})\cdot\dfrac{(1-m_{22}-m)}{3-m_{22}-m} \\ \frac{1}{2}(1-m_{22}-m) & m_{22} & m & \frac{1}{2}(1-m_{22}-m) \\ \frac{1}{2}(1-m_{22}-m) & m & m_{22} & \frac{1}{2}(1-m_{22}-m) \\ (1-m_{11})\cdot\dfrac{(1-m_{22}-m)}{3-m_{22}-m} & \dfrac{(1-m_{11})}{(3-m_{22}-m)} & \dfrac{1-m_{11}}{3-m_{22}-m} & m_{11} \end{matrix} \right\| \tag{3.23}$$

We proceed now to trace the evolution of our four-variety biopopulation. To compute their respective numbers at successive generations we must know what their initial numbers were. Suppose a single variety is initially present, say V_1 or 00. Call $N_0^1 = N$ its initial population. Then by (3.7) one generation later the average numbers are

$$\begin{aligned} N_1^1 &= 2(2m_{11} - 1)(1 - d_1)N \\ N_1^2 &= \frac{2(1 - m_{11})}{3 - m_{22} - m}(1 - d_1)N \\ N_1^3 &= N_1^2 \\ N_1^4 &= (1 - m_{22} - m)N_1^2. \end{aligned} \tag{3.24}$$

If $m_{11} \leq \frac{1}{2}$ or $d_1 = 1$, the first variety becomes extinct forthwith. And if $m_{11} = 1$ while $d < 1$, V_1 prospers without change. The *necessary and sufficient condition* for the evolutionary process to occur is then

$$\tfrac{1}{2} < m_{11} < 1 \quad \text{and} \quad 0 \leq d_1 < 1. \tag{3.25}$$

If this condition obtains then the lower bound of the population of varieties V_2 and V_3 at first generation is

$$\tfrac{2}{3}(1 - m_{11}) \cdot (1 - d_1)N.$$

Since this number equals at least 2, we infer that the lower bound of the stability of the parent variety is given by

$$m_{11} \geq 3/(1 - d_1)N. \tag{3.26}$$

Because the upper bound of m_{11} is 1, we infer that for a change of variety to occur it is necessary that

$$N > 3/1 - d_1. \tag{3.27}$$

For a low death rate of the parent variety, as few as four members of it could give rise to the differentiation process.

The above conditions concern only varieties V_2 and V_3. For variety V_4 to emerge a more severe condition must be met, as we gather from the last of equations (3.24), namely

$$(3.25) \ \& \ (m_{22} + m < 1) \quad \text{or, equivalently,} \quad (3.25) \ \& \ (m_{21} + m_{24} > 0).$$

Finally we write the general evolutionary equations (3.7) for an arbitrary generation (t value). Using the mutation matrix (3.23), we obtain the following system of first order linear difference equations:

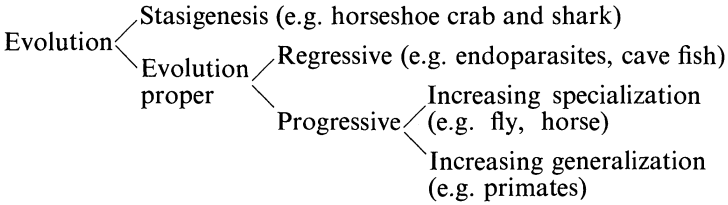
$$\begin{aligned} N_{t+1}^1 = {} & 2(2m_{11} - 1)(1 - d_1)N_t^1 + (1 - m_{22} - m)(1 - d_2)N_t^2 \\ & + (1 - m_{22} - m)(1 - d_3)N_t^3 \\ & + \frac{2(1 - m_{11})(1 - m_{22} - m)}{3 - m_{22} - m}(1 - d_4)N_t^4 \end{aligned}$$

$$\begin{aligned} N^2_{t+1} &= 2(2m_{22} - 1)(1 - d_2)\, N^2_t \\ &\quad + \frac{2(1 - m_{11})}{3 - m_{22} - m}(1 - d_1)N^1_t + 2m(1 - d_3)N^3_t \\ &\quad + \frac{2(1 - m_{11})}{3 - m_{22} - m}(1 - d_4)N^4_t \\ N^3_{t+1} &= 2(2m_{33} - 1)(1 - d_3)N^3_t + \frac{2(1 - m_{11})}{3 - m_{22} - m}(1 - d_1)N^1_t \\ &\quad + 2m(1 - d_2)N^2_t + \frac{2(1 - m_{11})}{3 - m_{22} - m}(1 - d_4)N^4_t \\ N^4_{t+1} &= 2(2m_{44} - 1)(1 - d_4)N^4_t \\ &\quad + \frac{2(1 - m_{11})(1 - m_{22} - m)}{3 - m_{22} - m}(1 - d_1)N^1_t \\ &\quad + (1 - m_{22} - m)(1 - d_2)N^2_t + (1 - m_{22} - m)(1 - d_3)N^3_t. \end{aligned}$$

This is the kind of linear algebra problem that computers love. But its solution is of no philosophical interest. What are of philosophical interest are the following points. First, the general evolution equations (3.7) can be turned, from a mere descriptive black box, into a mechanismic model representing the mutation-selection process – a model capable of making precise predictions about a certain evolutionary line. Second, this is a peculiarity not of our problem but of the synthetic theory of evolution, which is not a predictive theory unless it is adjoined precise data and hypotheses concerning the genetic makeup and the environment of the organisms concerned. (We shall come back to this problem in Vol. 6.)

3.6. *Bioprogress*

Evolutionary biologists agree that all modern complex organisms descend from primitive organisms: recall Figure 3.2. So, there would seem to be no doubt that evolution has been progressive. Not so fast. First of all, the following kinds of evolutionary process are usually distinguished:


Evolution
Stasigenesis (e.g. horseshoe crab and shark)
Evolution proper
Regressive (e.g. endoparasites, cave fish)
Progressive
Increasing specialization (e.g. fly, horse)
Increasing generalization (e.g. primates)

Some lineages, then, have advanced over certain periods while others have not. The second problem is that there is no agreement on the very notion of bioprogress – the more so since progress through increasing specialization is sure to lead to stagnation or extinction.

A number of alternative definitions of bioprogress have been proposed (cf. Williams, 1966, Bowler, 1976). Here are some: bioprogress is increasing histological differentiation, morphological complication, survivorship, adaptability, freedom from the environment, or amount of genetic information. All of these proposals presuppose that progress is global rather than partial (in particular respects), and none of them lends itself easily to quantitation. In particular, the information-theoretic definition is pseudoquantitative, as there is no unambiguous and objective way of measuring or computing the "quantity of information stored in an organism" (Lwoff, 1962; Apter and Wolpert, 1965). After all, biosystems are not identical with their genomes, and the latter are more than just information processing devices.

We shall take it that bioprogress consists in increasing biovalue, and shall proceed to refine the distinction (Ayala, 1974) between uniform (relentless) progress, on the one hand, and overall (net, average) progress on the other. Our elucidation will make use of the concept of system progeny (Definition 1. 15(v)) and Definition 3.9 of biovalue. Since the former was framed for systems in general, we shall first adapt it to biopopulations; and this adaptation will involve the notion of a variable collection, i.e. one with changing membership (Vol. 3, Definition 5.13). Let us then start with

DEFINITION 3.15 Let $\mathscr{P} = \{p_t \mid t \in T\}$ be a family of biopopulations of a certain kind, indexed by time (T). For any $p_0, p_t \in \mathscr{P}$, where $t > 0$,

(i) p_t *descends* from p_0, or $p_0 \prec p_t$, iff every member of p_0 descends from some member(s) of p_0;

(ii) the *progeny* of p_0 is the family of biopopulations that descend from p_0:

$$P(p_0) = \{p_t \in \mathscr{P} \mid p_0 \prec p_t\}.$$

We now add the idea that, if x and y are successive populations, x constitutes an improvement over y in a certain respect z if x performs better than y in doing z. More precisely, we propose

DEFINITION 3.16. Let p_0 be a biopopulation of organisms of a certain kind K, and call $P(p_0)$ the progeny of p_0. Further, call Σ a certain kind of subsystem (e.g. organ) of every member of p_t in $P(p_0)$, and assume that all the subsystems of that kind perform (roughly) the same function(s). Then

(i) the K's have *advanced on the whole* with regard to Σ between p_0 and p_t, for $t \gg 0$, iff the value of subsystem of kind Σ for the members of p_t is significantly greater than its value for the members of p_0;

(ii) the K's have *advanced uniformly* along $P(p_0)$ with regard to Σ between p_0 and p_t, for $t \gg 0$, iff the value of subsystem of kind Σ for the members of every population p_u in $P(p_0)$, between p_0 and p_t, is greater than its value for the immediate ancestors of the p_u's.

Whereas the first clause elucidates the notion of net or average progress, the second clarifies that of uniform progress. Uniform progress implies net progress but not conversely: i.e. there may be temporary setbacks in a progressive lineage,

Finally we make

DEFINITION 3.17 Under the nomenclature and assumptions of Definition 3.16, the K's have *progressed globally* between p_0 and p_t iff the value of every function of the members of p_t is greater than the value of the corresponding function for the members of p_0.

Remark 1 The above definitions do not discriminate between progress in a roughly constant environment, therefore at the cost of severe selection, and the accidental improvement of environmental conditions (or preadaptation). *Remark 2* Our definitions involve the concept of biovalue, hence that of goodness. Indeed, x is *good* for $y =_{df}$ the value of x for y is positive. However, these concepts are biological in our system, because biovalue was defined in strictly biological terms (Definition 3.9). Instead of injecting an anthropomorphic concept of value into biology, we have tried to derive an objective concept of value from the biological concept of health of a biosystem,

So much for definitions. Now is the time for commitment. Is there progress in evolution and, if so, is it universal and global or local and partial? According to some philosophers and biologists, from Aristotle to Lamarck, bioevolution is gradual and slow progress toward perfection. This hypothesis presupposes that the environment has remained constant for over 4 billion years, or that it has changed for the better of every species. This presupposition has been refuted by the evidence for continental drifts, glaciations, desertifications, earthquakes, and the depletion of a number of resources. While some such changes have been favorable to some species (those preadapted to the new environments), and even to the emergence of new species, they have been fatal to others. In any case most species have become extinct instead of persisting on the road of improvement.

Another presupposition of the hypothesis of universal and global progress is that there exists a fixed goal – an ideal or paradigm – towards which successive generations move. This hypothesis would make sense in a Platonic biology, where the archetype is perfect, every real organism being a sort of blurred carbon copy of the former. But of course there is no such thing as Platonic biology, although there are vestiges of Platonism in biotaxonomy. The very notion of an immaterial object – such as the Idea of the Perfect Mosquito – guiding the evolution of concrete things is inconsistent with the ontology of modern science. (Cf. Vol. 3, Introduction and Ch. 2.) Besides, since the environment changes rather erratically and sometimes dramatically, if there were a target then it should be moving rather than fixed, and so evolutionary processes zig-zag rather than go in a definite direction.

In sum, the biosphere does not obey the so-called "law of progress". Progress, when it *does* occur, is a trend not a law, and one occurring in *some* lineages, in *some* respects, and for *some* time. In the biosphere, just as in the sociosphere, there are not only progressives (e.g. mammals) but also conservatives (e.g. ants) and reactionaries (e.g. lice). Besides, whereas many progressives have become extinct, some conservatives are still around: drastic environmental changes and fierce competition can wreck the most brilliant careers.

For all the above reasons, the most we can assert is

POSTULATE 3.5 Every biospecies advances in some respects during some periods and in certain environments.

That is, bioprogress is often real but always partial, relative, and temporary. It is *partial* because an improvement in some respect(s) may constitute a drawback in other(s); e.g. an increase in size may confer greater strength but it also calls for greater energy consumption and less mobility. Progress is *relative* to the environment – moreover it can be active or passive: earned by mutation and selection, or won in the lottery of environmental cataclysms. And progress is *temporary* or limited to definite periods rather than eternal.

4. CONCLUDING REMARKS

The traditional biometaphysics are vitalism and mechanism. The strength of the former lies in its ability to ask questions, some of them pertinent, that the latter fails to answer. Whereas classical vitalism has hardly evolved, mechanism has. Two kinds of mechanism are popular today: (a) *physicochemicalism*, or the thesis that organisms are just physicochemical systems, and (b) *machinism*, or the thesis that organisms are machine-like systems, hence fashioned after a design or plan, and acting in a goal-seeking manner.

Physicochemicalism is the heir to traditional mechanism, and machinism a hybrid of mechanism and vitalism. Whereas the model of the first is the chemical reactor, that of the second is the computer. Neither of these biometaphysics is consistent with the whole of biology, although physicochemicalism has been fruitful where machinism has been sterile. Physicochemicalism is incompatible with biology because it denies emergence and levels, and machinism because it takes for granted and posits plans and goals where there are none.

The alternative we propose is *biosystemism*, or the application of systemism (Ch. 1, Sec. 4.2) to the study of life. Biosystemism recognizes the *bios* as an emergent level rooted to the chemical one. Therefore it accepts the lay difference between living and nonliving – denied, e.g. by Pirie (1960), Kendrew (1966) and Keosian (1968) – as well as the molecular-biological account of the self-assembly of biosystems from biochemical precursors, and the theory of evolution by genic changes and natural selection.

We are faced then with four different biometaphysics, each with its accompanying problem circle, method, and goal: see Table 3.1. The oldest and most persistent of them is vitalism, which is also the most popular biophilosophy. Actually there are two kinds of vitalism: ontological and methodological. The former states that organisms are distinguished by some immaterial principle, the latter that they are so complex and variable, that they cannot be tackled with the ordinary methods of science. Ontological vitalism has been pronounced dead long ago, even by biologists who continually lapse into it – such as Monod (1970) and Jacob (1970) when fantasizing about the *project téléonomique* of every organism. On the other hand the claim that organisms cannot be studied by the ordinary scientific method but call for a special method (i.e. methodological vitalism) is still very much alive except among molecular biologists (Bunge, 1973a, Ch.3).

As for machinism, its distinctive thesis is that organisms are like machines in their having a design or plan and a purpose. However, two kinds of machinism must be distinguished: verbal and serious. Verbal machinism restricts itself to repeating uncritically the organism-machine

TABLE 3.1

Four approaches to the study of life

		Mechanism		
	Vitalism	Machinism	Physico-chemicalism	Biosystemism
Ontological Framework	Life transcends matter and is guided by an immaterial principle	Organisms are machine-like: they have plans and goals.	Organisms are physicochemical systems.	Organisms are *sui generis* systems composed of biochemical systems and possessing emergent properties and laws.
Problem Circle	'What is it for?' questions.	'What is it for?' and 'How does it work?' questions.	'How is it composed?' and 'What are its properties?'	All biological problems, in particular regarding subsystems, their self-assembly, and evolution.
Method	Description of organism as a whole.	Physico-chemical methods plus building of analogs.		The whole array of particular methods of the various chapters of natural science.
Goals	Intuitive understanding of the whole	Technology-wise explanation and forecast.	Physicochemical explanation and prediction.	Biological explanation and prediction.

analogy and to warning that biology may learn more from engineering that from physics and chemistry (Polanyi, 1968). This kind of machinism can be discarded by noting that biology does not have a single law statement, let alone theory, containing the concepts of plan (or design) and purpose. (Even the most serious attempt to incorporate the concept of goal-directedness into biology, namely that of Sommerhoff (1974), fails to mathematize correctly that concept as well as to exhibit examples of laws and theories where the concept occurs.)

Serious contemporary machinism is the heir to Descartes and La Mettrie and culminated with the theory of self-reproducing automata (von Neumann, 1966). It is rather popular with the computer science crowd, as can be seen by scanning the *J. Computer and System Sciences.* This approach has stimulated, in particular, biological applications of the theories of automata and information. However, three decades of work along this line have yielded precious little; in particular, it has not contributed to the molecular biology revolution or to evolutionary biology. No wonder, for machinism skips the chemical level and ignores history: it thinks in terms of ready made systems rather than of enzymatic reactions or natural selection.

Machinism combines the negative aspect of mechanism (downward reductionism) whith the negative aspect of vitalism (upward reductionism). Its sole virtue is its clarity. Surely, machinism has stimulated the design and construction of ingenious mock organisms, such as Grey Walter's mechanical tortoise. However, since such artifacts are avowedly very different from the real thing they simulate, they do not teach us much about the latter. The only use of biosimulates is as toys or as slaves. In short, machinism is mistaken because it ignores that organisms are composed of chemical systems, rather than electromagnetic ones, and assigns them a property very few animals exhibit, namely that of seeking goals, and another that none have, namely optimal design or perfection.

As for physicochemicalism, it has undoubtedly been a fertile biophilosophy. In fact it has stimulated the birth of entire branches of science, such as biophysics, biochemistry, molecular biology, and bioengineering. In particular molecular biology might not have been born but for the hypothesis that chemical (in particular enzymatic) reactions constitute the root if not the marrow of life. In any case this conjecture sparked off, among other things, the chemical theory of heredity, which was the initial phase of molecular biology. (On the other hand physicalism, which tried to skip the chemical level, failed to contribute anything to the developments that culminated in the discovery of the double helix and the genetic "code": see Olby (1974). This is not to deny that physics, in particular X-ray crystallography, was indispensable. But it is one thing to use physical techniques and another to conceive of the organism as a physical system.)

Molecular biology, largely an offspring of the physicochemicalist branch of biometaphysics, has gone beyond digging up the chemical roots of life: it has shown that a number of properties that used to be regarded as typically biological – such as metabolism and self-reproduction – are actually biochemical. However, in their eagerness to rid biology of myth and explain it in terms of chemistry, molecular biologists have often stressed composition at the expense of organization or structure, as well as heredity to the detriment of environment. Consequently they have often overlooked the multi-level structure of the organism – organelle, cell, organ, multicellular organism. For this reason, although molecular biology is indispensable for explaining evolution, it is not sufficient: phenotypes and environments are just as important as genotypes. After

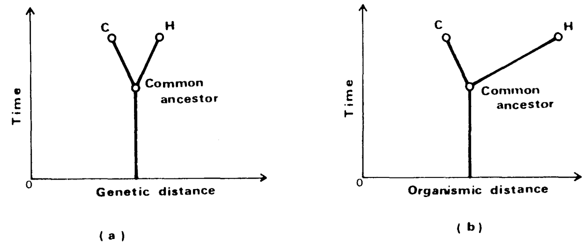


Fig. 3.7. Humans (H) and chimpanzees (C) are genetically about as different as two subspecies of mice (*a*). They differ mainly at the organ level (*b*), particularly in respect of the nervous system, hence of behavior. See King and Wilson (1975).

all, man and chimpanzee are hardly distinguishable at the biochemical level: see Figure 3.7.

To conclude. Vitalism is untenable but it has bequeathed us two important true principles:

V1 Life is emergent: it transcends the physico-chemical level.

V2 The biosphere has a level ("hierarchical") structure going from cell to organ to organism to population to ecosystem.

Physicochemicalism denies these principles but contributes another two important true hypotheses:

M1 Life emerged by the self-assembly of macromolecules.

M2 Some physical and chemical laws continue to operate in organisms.

Biosystemism adopts all four principles and adds its own:

S1 Organisms are systems composed of biochemical systems and possessing emergent properties.

S2 Every level in the biosphere has its own laws: laws of cells, of organisms, of populations, and of ecosystems.

S3 The units of biological study are the organism-in-its environment and its various subsystems (molecules, cells, organs) and supersystems (population, ecosystem, biosphere).

So much for the ontology of biosystemism. As for its methodology, it is epistemic reductionism, which is perfectly compatible with ontological pluralism (Bunge, 1977f), and boils down to

RULE 3.1. Each bioentity must be studied on its own level and should be explained with the help of its adjoining levels.

Thus an organ should not only be studied on its own level but should also be analyzed both downward (in terms of its subsystems) and upward (in terms of the entire organism in its environment). This holds particularly for the human brain, which is not only a neural system but also a subsystem of a component of social systems. But this is the story to be told next.

CHAPTER 4

MIND

The present chapter deals with one of the oldest, most intriguing, and most perverse of all the problems lying in the intersection of philosophy and science, namely the so-called mind-body problem. This is the system of questions about the nature of mind and its relations to matter.

Some eminent thinkers – notably Spencer (1862), E. du Bois-Reymond (1872), and Popper and Eccles (1977) – have pronounced the problem to be insoluble. I submit that part of the problem is that it is usually formulated in inadequate terms – namely those of ordinary language. These are inadequate not only because ordinary language is notoriously poor and inexact, to the point of being of little use in investigating the problem of matter, but also because it begs the question. Indeed European languages are loaded with a preconceived solution to the problem of mind, namely psychophysical dualism, or the doctrine that mind is something apart from matter. The very expression 'mind-body problem' suggests that mind and body are distinct entities on the same footing – i.e. things. We must then look to science for inspiration.

In science we do not speak of the motion-body problem, or of the reaction-substance problem, or of the digestion-digestive tract problem, or of the mobility-society problem. We speak instead of the motion *of* bodies, the reaction *of* chemicals, the digestive functions *of* the digestive tract, and the mobility *of* a society. We do not reify properties, states, or events – except when it comes to the properties, states, and events of the nervous system. We shall close this gap, that keeps the study of mind a scientific anomaly in the grip of myth, by replacing ordinary language by state space language, which is mathematically precise and is shared by science and scientific ontology.

Another difficulty besetting the so-called mind-body problem is the extreme youth of neuroscience: recall that Ramón y Cajal's neuronal hypothesis is less than one century old. Surely there were important sporadic discoveries in earlier times, starting with Hippocrates' hypothesis that the brain – rather than the heart or the liver – is the organ of emotion and thought. However, it is only in recent years that a concerted attack at all levels has been launched (Worden *et al.* (Eds.), 1975).

A third difficulty is that, far from being a strictly scientific problem, ours is also a philosophical one. In fact it is a traditional metaphysical problem and the one that has aroused the greatest passion – hence also the greatest caution. Moreover its very formulation presupposes a number of notions that are far from clear in traditional metaphysics, in particular those of substance, property, state, event, emergence, and level of organization. Thus any argument as to whether or not there is a mental substance, or whether there are mental states that fail to be brain states, calls for a prior elucidation of the general concepts of substance and state – an elucidation that most philosophers do not care to perform.

Fourth and last, another reason for the backward state of research into the so-called mind-body problem is that it belongs not only to science and to philosophy but also to ideology. Indeed all religions and some political ideologies have a vested interest in it and some even claim exclusive ownership. So, far from being interested in the investigation of the problem, they are anxious to have us accept their own ready-made solution.

In short, there have been many obstacles in the way of a scientific study of the problem. Fortunately some of these obstacles are disappearing rather quickly, and the science of mind – i.e. psychology – is making great strides. Three formerly separate currents are converging on it: neuroscience, psychology, and philosophy. Indeed, over the past few decades neurophysiologists and neurologists have begun to study the mental functions of neural systems, physiological psychologists have started to disclose the neural mechanisms "mediating" behavior, and philosophers have begun to analyze some of the key concepts involved in the question, and to overhaul the ancient doctrine that mental functions are a kind of bodily functions. There is still a long way to go before maturity is attained, but the science of mind has finally made a start.

This chapter, like the others in this book, will offer only a general framework – but, of course, one consistent with the basic ideas of system, property, law, and change prevailing in science and expounded earlier in the work.

1. CENTRAL NERVOUS SYSTEM

1.1. *Philosophical Background*

Perceiving, feeling, remembering, imagining, willing, and thinking are usually said to be mental states or processes. Since in science and in our ontology there are no states or processes in themselves, but only states *of* some entity and processes *in* some thing, we must ask *what* minds – i.e. what is the thing that perceives, feels, remembers, imagines, wills, and thinks. This is the very core of the so-called mind-body problem, i.e. the identification of the subject or referent of the mentalist predicates.

Those who hope the mind-body problem to be soluble have proposed two solutions to it: that what does the minding (perceiving, feeling, thinking, etc.) is the mind (or soul or spirit), and that it is the brain. According to the former the mind is an immaterial entity wherein all mental states and processes occur: feelings, memories, ideas, and the like would be in or of the mind, which would in turn be an entity separate from the body. According to the second answer the mind is not a separate entity but a set of brain functions: perceiving, imagining, thinking, and the like would be brain processes. The former thesis is called *psychophysical dualism*, the latter *materialism* or *naturalism*.

However, there are a number of varieties of dualism and of materialism. (Cf. Vesey, 1964; Smythies, 1965; Armstrong, 1968; O'Connor, 1969; Borst, 1970; Popper and Eccles, 1977; Margolis, 1978.). Our main task in this chapter is not to discuss them but to formulate a version of materialism consistent with the principles laid down in the previous chapters and in Vol. 3 of this work. This is *emergentist materialism*, or the doctrine according to which (*a*) all mental states, events and processes are states of, or events and processes in, the brains of vertebrates; (*b*) these states, events and processes are emergent relative to those of the cellular components of the brain, and (*c*) the so-called psychophysical (or psychosomatic) relations are interactions between different subsystems of the brain, or among them and other components of the organism (Bunge, 1977e). This kind of materialism is monistic with respect to substance and pluralistic with respect to properties.

Here are some reasons for favoring emergentist materialism:

(i) because it eschews the ghostly mental substance without thereby denying mental facts, emergentist materialism is *compatible with the scientific approach*;

(ii) emergentist materialism – as we hope to show – is *free from the fuzziness that characterizes dualism* with its talk of an immaterial mind and of "correlations" between it and the body;

(iii) unlike dualism, emergentist materialism is *consistent with the general concepts of state and event that can be gleaned from all the sciences*. (On the other hand according to dualism mental states would be the *only* states that fail to be states of some thing, and mental events the *only* events that fail to be changes of state of some thing – this being why dualism agrees more with theology than with science.)

(iv) Unlike dualism, emergentist materialism *fosters interaction between psychology and the other sciences*, in particular neuroscience, and this precisely because it regards mental events as special biological events;

(v) unlike dualism, which digs an unbridgeable chasm between man and beast, emergentist materialism *accords with evolutionary biology*, which – by uncovering the gradual development of mental faculties along certain lineages – refutes the complacent superstition that only Man is endowed with Mind;

(vi) unlike dualism, which postulates an unchanging mind, emergentist materialism *agrees with developmental psychology* and neurophysiology, which exhibit the gradual maturation of the brain and behavior.

The only drawback to emergentist materialism is its so far not having been formulated in a precise and detailed way. Indeed, emergentist materialism is little more than a programmatic hypothesis with enormous heuristic power. We shall try and remedy this shortcoming in the course of this chapter, by building a general theory of the mind consistent not only with our ontological principles but also with current neuroscience and psychology.

1.2. *Neural Units*

We shall deal throughout this chapter with systems of a special kind, namely organisms endowed with nervous systems. And we shall pay special attention to animals endowed with a central nervous system, or CNS for short. Nervous systems come in a number of degrees of complexity, from primitive worms with half a dozen neurons, to insects with about 100,000, to human with nearly a hundred billion (i.e. 10^{11}). The corresponding behavior levels range from automatic to highly creative. And the varieties of inner life from nil to extremely rich. Since the lower forms of nervous system functioning and animal behavior are generally recognized to be explainable along strictly biological lines, we shall concentrate on the higher functions – i.e. on those which, according to popular belief, require the presence of a soul, spirit, or mind inaccessible to science.

The following levels may be distinguished in a CNS:

Subcellular: neuron's membranes, synaptic boutons, dendrites.

Cellular: neurons and glial cells.

Neural microsystems: multineuronal systems (fixed or itinerant assemblies of thousands or millions of neurons in primates).

Neural macrosystems: systems of thousands of neural microsystems. In primates the following sublevels are distinguished:

> Uncommitted cortex, somethetic system, visual and hearing systems, temperature regulating system, etc.;
> cerebral hemispheres;
> brain;
> central nervous system (CNS), or brain *cum* spinal cord;
> neuroendocrine system (NES), or nervous system *cum* endocrine glands (e.g. pituitary and adrenals).

Every mental process presumably engages millions of neurons and occurs not only at the neural system level but also at the cellular and subcellular levels. Moreover, given the strong coupling between the CNS and the rest of the body, every mental act is bound to engage also some non-nervous tissue. E.g., voluntary movement is an activity not only of certain neural systems but also of certain muscles and endocrine glands. Thus the various neural systems (micro and macro) can be distinguished from one another but they cannot be detached from each other or from the rest of the body, They do the minding but not in isolation from other subsystems of the body.

While some neural systems have permanent components others don't. Thus the visual system has a constant composition, i.e. is composed of anatomically identifiable subsystems – the eyes, the optic nerves, the optic tracts, the lateral geniculate nuclei, and the visual cortex. Other neural systems may not have constant components and so may not be carved out by the surgeon: far from having a constant location, they may be itinerant neuron assemblies formed just for the occasion (Craik, 1966; Bindra, 1976). See Figure 4.1.

We shall call a *psychon* any neural unit capable of discharging mental functions of some kind. There are three main views concerning the nature and size of psychons: neuronism, holism, and systemism. According to neuronism single neurons may possess certain mental abilities – e.g. that of issuing commands. There is no evidence for this view. Holists, on the other hand, assume that nothing short of the whole brain can mind: they speak of 'mass action' and favor holographic theories of memory and other mental functions. Although there is plenty of evidence for the strong coupling among a number of neural systems, there is also evidence for localization—of, e.g., pleasure and speech.

So we are left with *psychosystemism*, or the hypothesis that the brain,

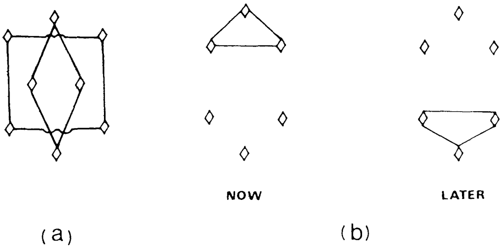


Fig. 4.1. (*a*) Two neural microsystems occupying roughly the same spatial region yet functionally detached: they 'subserve' different functions. (*b*) Psychons can be itinerant even if the components stay in place, because their connections change in time.

far from being either a heap of self-sufficient units or a homogeneous jelly, is a *system of specialized subsystems* or organs, some of which are itinerant rather than fixed. There is multiple evidence for psychosystemism, so we shall adopt it. Besides, psychosystemism accords with our systemist world view and furthermore it suggests the most fruitful research strategy, namely that of investigating the brain as a whole as well as all of its subsystems, their mutual couplings, and their links with non-neural subsystems, particularly the endocrine system.

The systemic approach to the CNS is necessary but insufficient: we must also take development and evolution into account. Dualism regards the mind as an invariable entity, holding that each person keeps his or her "personal identity" throughout life. However, the empirical evidence is against this thesis: the newborn human has hardly a mind, after a few months has learned a number of tricks, and continues to learn until senility sets in. In short the CNS of vertebrates is plastic. Moreover it has evolved from modest beginnings. The believers in the myth of the immaterial mind must ignore the evidence that the mind has evolved along with other bodily functions. So, we shall ignore the dualistic myth of personal identity: we shall hold instead that the only constant thing about John Smith is his name.

1.3. *Initial Definitions*

We start by making

DEFINITION 4.1 An *animal* is an organism that
(i) feeds on other biosystems;

(ii) can move as a whole toward or away from some stimuli;

(iii) can explore its surroundings in search of food, water, or stimulation, or to avoid stimuli that might impair its health,

We shall next define a nervous system as an information biosystem in an animal. We can and must do without the concept of information in biochemistry and in cellular biology, not however when dealing with organisms endowed with endocrine or, *a fortiori*, neuroendocrine systems. Here the environmental and the internal inputs not only are nonspecific energy transfers but sometimes also generate specific signals conveying definite information about events. Indeed there is a definite code – a naturally evolved one of course – whereby stimuli of a certain kind (e.g. thermal) are "translated" (transduced) into neural pulses in the organism; and these signals in turn activate other components, e.g. muscles or glands. These are signals proper not just neutral energy fluxes: they carry information and they trigger the release of much greater quantities of energy than they transport. The nervous system, in sum, is a genuine information network, for it detects, generates, and transmits information in accordance with definite natural codes. Hence we make

DEFINITION 4.2 A system is a *nervous system* iff it is an information biosystem (i.e. a living signal-detecting-and-processing system) such that

(i) it is composed of (living) cells;

(ii) it is or has been a proper part of a multicellular animal (i.e. the complement of the latter in the immediate environment of the system);

(iii) its structure includes (*a*) the regulation or control of some of the biofunctions of the animal, and (*b*) the detection of internal and environmental events, as well as the transmission of signals triggered by such events.

DEFINITION 4.3. A biosystem is a *neural* (or *neuronal*) system iff it is a subsystem of a nervous system.

DEFINITION 4.4. A biosystem is a *neuron* iff it is a cellular component of a neural system.

Remark 1 The vertebrate nervous system has a third structural property: it interacts directly with every other system of the animal. And the CNSs of higher vertebrates are more than information processors: they are also information generators. More on this in Sec. 4.5. *Remark 2* Neural circuits and ganglions, cell assemblies (Hebb, 1949) and neuron populations

(Freeman, 1973), working constellations (Luria, 1966) and itinerant neural systems or pexgos (Bindra, 1976) are neural systems, and so are the functional columns of cortical neurons (Mountcastle, 1957; Powell and Mountcastle, 1959; Hubel and Wiesel, 1963). *Remark 3* The neural systems capable of perceiving signals, imagining, or willing, are probably composed of millions or billions of neurons. However, even neuron systems composed of a few neurons are bound to possess properties that its components lack, i.e. emergent properties. Therefore neuronal systems are ontologically irreducible entities even if they are (up to a point) epistemically reducible, i.e. explainable in terms of composition and structure. *Remark 4* The preceding remark counters the dualistic argument that, because all cortical neurons look alike (to the untrained eye), they cannot be attributed the diversity of functions imputed to them by the monists. Actually no two neurons are alike. And even if they were, we know that they form systems (in particular, columns) – and, as any child experienced with Lego or Meccano knows, there is no end to the variety of systems that can be formed with a few elements. *Remark 5* Exclusive attention to the individual neuron and its alleged "all or none" firing pattern is misleading for it favors focusing on discontinuous processes and analogies with digital computers. Mental processes involve presumably millions of neurons and therefore continuous or quasi-continuous processes. This allows one to build continuous models of neural systems (Rashevsky, 1972; MacGregor and Lewis, 1977). And it disposes of the so-called "grain objection" to the psychoneural identity hypothesis, namely that whereas brain events are spatially and temporally discrete, we can have experience of continuous expanses of red.

The various neuronal systems differ in their mode of connection. There are several types of interneuronal and intersystem connection: constant or fixed, variable in a regular way, and variable in a random way. In order to discuss neural connectivity we had better define this notion. One way of doing it is as follows. Call $C_t(a, b)$ the strength of the connection from neuron (or neuron assembly) a to neuron (or neuron assembly) b of a given system at time t. In general, $C_t(a, b) \neq C_t(b, a)$. The $a - b$ connection is *excitatory* at t iff $C_t(a, b)$ is positive, and *inhibitory* iff $C_t(a, b)$ is negative. The excitation or inhibition caused by cell a on target cell b equals the output of a times $C_t(a, b)$. And the total input to cell b is obtained by adding all the partial inputs (i.e. by summing over a). The matrix $\|C_t(a, b)\|$ displays the total connectivity of the system at time t. The connectivity of a neural system with 1 million neurons is represented by a 1,000,000 × 1,000,000 matrix (lumping the hundreds of synapses of each cell into a single junction). Since we shall make heavy use of this concept we may as well display it for future reference:

DEFINITION 4.5. Let ν be a neural system and $\mathscr{C}_t(\nu)$ the neuronal composition of ν at time t. Further, call

$$C_t\colon \mathscr{C}_t(\nu) \times \mathscr{C}_t(\nu) \longrightarrow [-1, 1]$$

the function such that $C_t(a, b)$, for $a, b \in \mathscr{C}_t(\nu)$, is the strength (intensity) of the connection (coupling, link) from neuron a to neuron b at time t. Then the *connectivity* of ν at t is represented by the matrix formed by all the connection values, i.e.

$$\mathbb{C}_t = \|C_t(a, b)\|.$$

DEFINITION 4.6. A connectivity is *constant* iff it does not change once established (i.e. iff $\mathbb{C}_t$ is independent of time). Otherwise it is *variable*.

DEFINITION 4.7. A neuronal system is *plastic* (or *uncommitted*, or *modifiable*, or *self-organizable*) iff its connectivity is variable throughout the animal's life. Otherwise (i.e. if it is constant from birth or from a certain stage in the development of the animal), the system is *committed* (or *wired-in*, or *prewired*, or *preprogrammed*).

DEFINITION 4.8. Every plastic neural system is called a *psychon*.

So much for our preliminary definitions. We are now ready for the first bunch of axioms of our theory of the mind.

1.4. *Basic Assumptions*

Our first hypothesis is rather obvious:

POSTULATE 4.1. All animals with a nervous system have neuronal systems that are committed, and some animals have also neuronal systems that are plastic (uncommitted, self-organizable).

Apparently worms, insects and other lowly animals have only totally (or nearly totally) wired-in neuronal systems. On the other hand the plastic cortex of man is the largest in all animals and a sort of vast reserve army ready to meet emergencies, most of which never happen. Hence whereas the behavioral repertoire of lowly animals is predetermined (programmed), that of man (and many another vertebrate) can evolve in the course of his life. More on behavioral plasticity in Sec. 3.3.

POSTULATE 4.2. The neuronal systems that regulate (control) the internal milieu, as well as all the biofunctions of the newborn animal, are committed (wired-in).

Example The temperature and acidity of the intercellular fluid are regulated mainly by certain wired-in neuronal systems, and so are the breathing and sucking movements of the newborn animal. But they are not the only regulators: also the endocrine system plays an important role in internal control.

POSTULATE 4.3 The plastic (uncommitted) neuronal systems of an animal (i.e. its psychons) are coupled to form a supersystem, namely the *plastic neural supersystem* of the animal. Symbol: P.

POSTULATE 4.4 Every animal endowed with psychons (plastic neuronal systems) is capable of acquiring new biofunctions in the course of its life.

From Postulates 4.1 and 4.4 it follows that some animals are in fact capable of acquiring new biofunctions in the course of their lives. (But of course the emergence of new biofunctions is only an indicator, not a proof, of the existence of plastic neuronal systems.)

Finally we make

DEFINITION 4.9 Every neural function involving a psychon (or plastic neural system) with a regular connectivity (i.e. one that is constant or else varies regularly), is said to be *learned.*

In other words, we assume learning to consist in the formation of new neural systems, i.e. in establishing permanent connections among neurons or facilitating ephemeral (but repeatable) neuron interconnections. Some such connections may be formed accidentally (at random); if valuable, they have a chance of becoming established or of recurring. These ideas go back to Ramón y Cajal and Hebb: they proposed the use-disuse hypothesis, which has been the object of much recent experiment (e.g. Moore, 1976; Rutledge, 1976) and theory (e.g. Malsburg, 1973; Cowan, 1976).

Let us proceed to refine the above ideas with the help of the concepts of state function and state space (Ch. 1, Sec. 2.2), and to introduce the special concept of mental function.

2. BRAIN STATES

2.1 *Brain Functions*

We have defined the functions of a system as the set of processes in the system, i.e. what the system does; and the specific functions of the system as those that the system, but no thing of any other kind, is capable of doing (Ch.3, Sec. 2.2, Def. 3.8). Just as there is no system without functions (processes), so there is no function without a system that does the functioning. When a neuroscientist says that function *X* is *mediated* or *subserved* by "structure" (neuron system) *Y*, he means that *Y does X* and, more precisely, that *X* is among the specific activities or processes of *Y*,

Being a multibillion-neuron system, the mammalian CNS engages in activities of a large variety, from synthesizing proteins to controlling the functions of other organs to performing higher functions such as forming a map of the surroundings. We shall not be concerned with the nonspecific or "household" CNS functions, except to note that, if anything, they are more intense than anywhere else in the body, as judged from the rates of oxygen intake and protein turnover. We shall instead focus on the specific functions of the CNS, which may be grouped into two large classes: control and cognition.

Control can be of internal systems (e.g. digestion), of external moving parts (motor), or of sensory inflow. And cognition can be of external events, of internal activity except the brain's, or of brain activity itself – in which case it is called 'consciousness'. We know from Chapter 3 that control is inherent in every biosystem and, indeed, in every biochemical system. What is peculiar to the control functions of the CNS is that they are ultimate and integrative, i.e. that there is no appealing them and that they concern the whole organism. But we shall not be concerned with the control functions of the CNS because they are nowadays of little philosophical interest. We shall instead deal with the cognitive activities or functions of the CNS.

In the first place there is the cognition of the external world, which can be direct (perceptual) or indirect (intellectual). The former shall concern us in Sec. 3.1, the latter in Sec. 4.3. Next come the various kinds of information that the brain receives from the rest of the body, and integrates and processes. Some of them, like the visual images, are rather specific and localized; others, like elation and depression, are rather nonspecific and nonlocalized. But all of them are functionally related to further bodily events: they are signalled by physiological changes outside the brain and in turn react upon them. Finally comes the brain's cognition of some of its own states, i.e. self-consciousness. We shall deal with it in Sec. 5.1.

We proceed now to explain the state space representation of CNS activity. Consider an arbitrary neuronal system, be it a small neuronal circuit (fixed or itinerant), a sizable subsystem of the CNS, or the entire brain. Like every other system, it can be represented by a state function $\mathbb{F} = \langle F_1, F_2, \ldots, F_n \rangle$. If we are interested not in the precise localization of the system and its components, but only in its overall nonspatial properties, we can assume that $\mathbb{F}$ is only a time-dependent function whose values are n-tuples of real numbers. I.e. we can set $\mathbb{F} : T \rightarrow \mathbb{R}^n$, with $T \subseteq \mathbb{R}$, and $\mathbb{R}$ the set of real numbers. (We take the host of the neural system, i.e. the whole animal, as its reference frame.)

Each component F_i of the state function $\mathbb{F}$ can be decomposed into a constant (or nearly constant) part F_i^c and a variable part F_i^v: see Figure 4.2. Obviously, either can be zero during the period of interest. However, the point is that, while the rate of change of F_i^c vanishes at all times (i.e. $\dot{F}_i^c = 0$), that of F_i^v does not. The latter may therefore be taken to represent the activity of the neuronal system in the ith respect. For this reason we make.

DEFINITION 4.10 Let $\mathbb{F} : T \rightarrow \mathbb{R}^n$ be a state function for a neural system ν, and let $\mathbb{F} = \mathbb{F}^c + \mathbb{F}^v$, with $\dot{\mathbb{F}}^c = 0$ (zero rate of change) for all $t \in T$. Then

(i) ν is *active* at time t iff $\mathbb{F}^v(t) \neq 0$;

(ii) the *intensity of the activity* of ν over the time lapse $\tau \subset T$ equals the fraction of components of ν active during τ;

(iii) the *state of activity* of ν at time t is $s = \mathbb{F}^v(t)$;

(iv) the (total) *process* (or *function*) ν is engaged in over the time interval $\tau \subset T$ is the set of states of activity of ν:

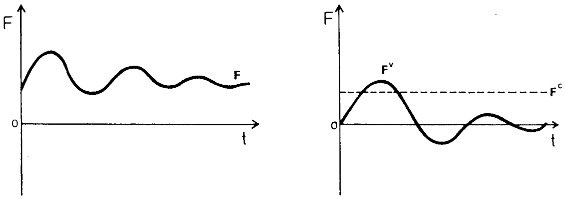


Fig. 4.2. Decomposition of a state function $\mathbb{F}$ into a constant part $\mathbb{F}^c$ and a variable part $\mathbb{F}^v$.

$$\pi(\nu, \tau) = \{\mathrm{F}^{\nu}(t) \mid t \in \tau\}.$$

Whereas single neurons may pass twice through almost the same state of activity, it is most unlikely that a neural system composed of a million neurons will do the same. We may assume that there are no two identical successive states of activity of a medium-sized neural system and, *a fortiori*, of a gross neural system such as, e.g., the hippocampus. Since mental states will be identified with some such states of activity, it follows that no mental event ever recurs in exactly the same form.

We now introduce the concept of specific activity of a neural system as a mere application (specification) of the concept of specific function of a biosystem (Definition 3.8 (ii)):

DEFINITION 4.11 Let $\pi(\nu, \tau)$ be the total process (or function) of a neural system ν in an animal b during the time interval $\tau \subset T$. The corresponding *specific function* (*activity, process*) of ν during τ is

$$\pi_s(\nu, \tau) = \pi(\nu, \tau) - \bigcup_{\mu \lessdot b} \pi(\mu, \tau) \quad \text{with} \quad \mu \neq \nu.$$

We have postulated that every thing changes spontaneously in some respects, inductively in others (Vol. 3, Postulate 5.11). In particular, neurons are never quiescent, and the human CNS is active even when sensory stimulation stops. We shall therefore adopt the hypothesis that the CNS, and every neural subsystem of it, is constantly active even in the absence of external stimuli. To formulate this principle consider m neurons synapsing on to a given neuron i. The rate of change of the latter's state of activity can be assumed to consist of an autonomous (spontaneous) part and another proportional to the signals of the afferent fibers:

$$\dot{F}_i^{\nu}(t) = A_i(t) + \sum_{j=1}^{m} C_t(j, i)\, F_j^{\nu}(t),$$

where $C_t(j, i)$ is the strength of the $j - i$ connection (cf. Definition 4.5). This system of equations is incomplete. It must be supplemented with a system of equations for the rates of change of the connectivities. A popular hypothesis is

$$C_t(i, j) = c_{ij} \cdot \exp(-\,|x|/b_{ij}),$$

where the c_{ij} and b_{ij} are real numbers, and $|x|$ the distance between the units i and j. Another interesting candidate would seem to be

$$\dot{C}_t(i, j) = c_{ij}F_i^v.$$

(For some mathematical models of the dynamics of cortical systems, see Marr (1970), Wilson and Cowan (1973) and Nass and Cooper (1975).)

However, here we are interested only in the general principle:

POSTULATE 4.5 For any neural system ν of an animal, the instantaneous state of activity of ν decomposes additively into two functions: $\mathbb{F}^v = \mathbb{A} + \mathbb{E}$, where $\mathbb{A}$ does not vanish for all $t \in T$, whereas $\mathbb{E}$ depends upon the actions of other subsystems of the animal upon ν.

DEFINTION 4.12 Let $\mathbb{F}^v = \mathbb{A} + \mathbb{E}$ be the active part of the state function of a neural system ν. Then $\mathbb{A}(t)$ is the *state of spontaneous activity* of ν at time t, and $\mathbb{E}(t)$ the *state of induced* (or *stimulated*) *activity* of ν at t.

Once we know, or pretend to know, how a neural system works, we may attempt to find out how two or more neural systems combine to form a supersystem. Suppose ν_1 and ν_2 are two neural systems of a given animal, with the same number m of neurons each. Call $\mathbb{F}_1^v$ and $\mathbb{F}_2^v$ their respective active state functions. And suppose further that the neurons in ν_1 synapse on to those of ν_2, so that ν_2 has a stimulated or induced activity in addition to its spontaneous activity. The simplest assumption is of course that the induced activity $\mathbb{E}_2$ of ν_2 depends linearly upon $\mathbb{F}_1^v$, i.e. $\mathbb{E}_2 = C \cdot \mathbb{F}_1^v$, where C is an $m \times m$ matrix representing the inter-system connectivity. This connectivity is time-dependent. We assume that it strengthens with use and weakens with disuse.

If ν_1 fails to activate ν_2, the value of C will have decreased, after one time interval, to $C(t + 1) = a\ C(t)$, with $0 < a < 1$. After k time intervals, $C(t + k) = a^k\ C(t)$, which approaches zero as time goes by. If on the other hand ν_2 is activated by ν_1, the connectivity will be strengthened in proportion to the simultaneous activity of the two neural systems. I.e. the increment in connectivity will be $\Delta C(t) = b\mathbb{F}_1^v \cdot \mathbb{F}_2^v$, where b is an $m \times m$ matrix of real numbers, some positive, others negative. Hence in general the connectivity value at time t will be

$$C(t) = a\ C(t - 1) + b\ \mathbb{F}_1^v(t) \cdot \mathbb{F}_2^v(t),$$

so that the state of activity of ν_2 at t will be represented by

$$\mathbb{F}_2^v(t) = \mathbb{A}_2(t) + aC(t - 1) \cdot \mathbb{F}_1^v(t) + b(\mathbb{F}_1^v(t) \cdot \mathbb{F}_2^v(t)) \cdot \mathbb{F}_1^v(t).$$

This is sheer speculation, though one in line with current trends in neural modeling. Its philosophical interest lies in showing that it is possible to construe mental states and events as states and events of neural systems, hence to treat them mathematically. More on this anon.

2.2. *Mental States and Processes*

Every fact experienced introspectively as mental is identical with some brain activity: this, in a nutshell, is the neurobiological or materialist hypothesis of the mind. For example, vision consists in the activity of neural systems in the visual system; learning is the formation of new neural connections; and memory is the activation of neural connections.

All mental activity is brain activity but the converse is false. For example some of the activities of the cortical component of the auditory system are mental; those of the committed or wired-in components (e.g. the ear drum and the cochlea) are not mental. Only plastic neural systems are capable of learning and they are the "seats" or "neural correlates" of the mental. More precisely, we assume that the mental is the specific function of certain plastic neuronal systems. Our assumption takes the form of

DEFINITION 4.13 Let b be an animal endowed with a plastic neural system P. Then

(i) b undergoes a *mental process* (or performs a mental function) during the time interval τ iff P has a subsystem ν such that ν is engaged in a specific process during τ;

(ii) every state (or stage) in a mental process of b is a *mental state* of b.

For example, acts of will are presumably specific activities of neuron modules in the forebrain. On the other hand hunger, thirst, fear, rage, and sexual urge are processes in subcortical systems (mainly hypothalamic and limbic), hence are nonmental according to our definition. What is a mental process is the consciousness of any such states – consciousness which is a process in some subsystem of P.

The following are immediate consequences of four previous definition:

COROLLARY 4.1. All and only animals endowed with plastic neural systems are capable of being in mental states (or undergoing mental processes).

COROLLARY 4.2 All mental disorders are neural disorders.

Consequently, the division between neurology and psychiatry, suggested by psychophysical dualism, is unreasonable.

COROLLARY 4.3 Mental functions (processes) cease with the death of the corresponding neural systems.

Examples Brain injuries, strokes, and of course death bring about the cessation of the mental functions normally discharged by the affected neural systems. Dualism does not explain such a "correlation".

COROLLARY 4.4 Mental functions (processes) cannot be directly transferred (i.e. without any physical channels) from one brain to another.

Consequently extrasensory perception is out of the question. (The Montreal magician Henry Morgan suggests that it should be renamed 'extrasensory deception'.)

We come now to the central concept in the philosophy of mind:

DEFINITION 4.14. Let P be the plastic (uncommitted) supersystem of an animal b of species K. Then

(i) the *mind* of b during the period τ is the union of all the mental processes (functions) that components of P engage in during τ:

$$m(b, \tau) = \bigcup_{x \prec P} \pi_s(x, \tau);$$

(ii) the *K-mind*, or *mind of species K*, during period τ, is the union of the minds of its members during τ:

$$M(K, \tau) = \bigcup_{x \in K} m(x, \tau).$$

Since the members of the set called 'mind' are brain functions (processes), it makes no sense to say that the brain is the physical "basis" of the mind. And since the human mind is nothing but the union of all the individual human minds, it makes no sense to speak of the collective mind of mankind as if it were an entity or even a functional system. On the other hand the mind of an individual animal does have functional unity: it is a functional system (Definition 1.9 in Ch. 1, Sec. 1.7), for every process (function) in it is lawfully related to at least one other process (function) in it. The "neural basis" of the unity of the mind, on which many philosophers have insisted, is this: the components of the plastic neural supersystem P, far from being uncoupled, form a system by Postulate 4.3. Hence

COROLLARY 4.5. The mental functions of (processes in) the plastic neural supersystem of an animal are coupled to one another, i.e. they form a functional system. (In short, for every animal b and every period τ of its existence, $m(b, \tau)$, if nonvoid, is a functional system.)

Consequently when a neurosurgeon cuts a living primate brain into two, he splits the plastic system into two, and thus also the corresponding mind of the animal. (See Corollary 4.7.)

Our last convention in this section is

DEFINITION 4.15. Let x be an object and b an animal endowed with a plastic neural system. Then

(i) x *is in the mind* of b iff x is a mental state or a mental process of b (i.e. if b has a plastic neural system engaged in a specific process containing x as a member or as a subset);

(ii) x *is in the mind*(or *is mental*) iff there is at least one animal y such that x is in the mind of y.

Remark 1 A mental state, or state of mind, such as one of persistent pain, or of perceiving a stationary thing, may be roughly constant. However, any such mental state is a state of activity of some neural system: no such activity, no mentation. In other words, mental states are more like states of motion than like static states such as the state of a gas in a rigid enclosure. *Remark 2* We have stated that mental states are neural states (or rather processes), not that every neural state has a mental "correlate". Although different mental states are different neural states (or rather processes), the converse is not true: a given mental state could be a process now in one itinerant neural system, now in another. Surely there are bound to be differences: even when doing routine sums, for all we know we may engage a different itinerant neural system each time – but we won't notice the difference introspectively (Bindra, 1976). (That is, calling N the set of neural states that happen to be mental, and M the set of mental states distinguishable introspectively, $M = N/\sim$, where $\sim$ is the relation of introspective equivalence.) *Remark 3* We have not characterized mental states independently of brain states, as dualists would like us to do. We have not done so for the following reasons: (*a*) mentalist predicates, such as 'sees red' and 'thinks hard', though indispensable, are coarse and vulgar rather than exact and scientific; (*b*) the whole point of the neurophysiological approach to mind is to get away from the ordinary knowledge approach and render mind accessible to science; (*c*) if mental events are characterized independently of brain events, then the identity theory turns out to be idle or false (Brandt and Kim, 1967). *Remark 4* The dualist claim that the mental must be nonphysical because it cannot be described with the help of physical concepts, such as those of pressure and conductivity, misses the point. Firstly, although such events as nightmares and feeling a sour taste are not describable in strictly physical terms, they are, at least in principle, describable in neurophysiological terms. Secondly, in ordinary life we may never care to describe mental states in neural terms, just as we do not care to describe such ordinary physical events as a car crash or the burning of a light bulb with the paraphernalia of physical theory. It would not be worth our while to do so; we are satisfied knowing that, if need be, such events, or some of their features, could be accounted for in scientific terms. *Remark 5 Pace* Armstrong (1968), mind and brain are not identical: there is no more brain-mind identity than there is lung-respiration identity. In our version of the so-called identity theory, the set of mental states is a subset of the states of the plastic neural system of the animal: see Figure 4.3.

2.3. *Psychosomatic Interaction*

The interactionist variety of dualism, from Descartes (1649) to Popper and Eccles (1977), holds that matter and mind, though heterogeneous and separate, act upon one another. This doctrine is untenable if only because

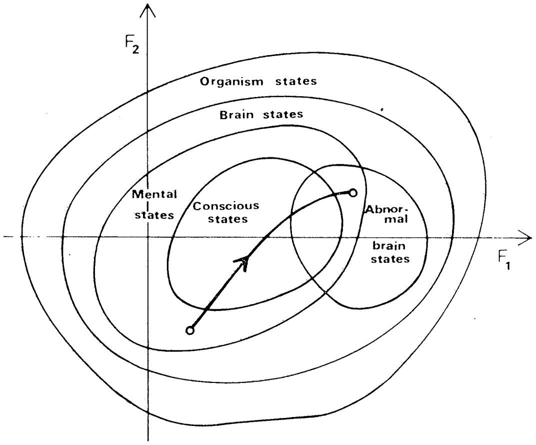


Fig. 4.3. Projection of the state space of a human being. Both axes represent neurophysiological properties; one of them is identical with a psychological property. The arc of curve represents a mental process, such as composing a melody, which is only partly conscious. One part of it is abnormal, e.g. consists of auditory hallucinations.

interactions are well defined only for concrete things (Vol. 3, Ch. 5, Def. 5.30). To speak of mind-body interaction is just as mistaken as to speak of shape-body, or property-thing, or composition-system, or behavior-animal interactions. However, it does make sense to speak of mental-bodily interactions provided this expression is taken to abbreviate "interactions among plastic (uncommitted) neural systems, on the one hand, and committed neural systems or bodily systems that are not part of the CNS, on the other". *Example* The interactions between cortical and subcortical regions of the CNS, between sensory and motor areas, between ideational neural systems and external receptors, between the visual cortex and the pituitary, etc.

Paradoxically enough, while the dualist has no right to claim that mental events influence nonmental bodily events (because he has no clear idea of such influence), the monist is entitled to this opinion. Indeed, since mental events are for him neural events of a kind, they can influence and even cause other events in any subsystem of the same body because of the integrative action of the CNS. In short, because mental events are neural events, and the causal relation is defined for pairs of events in concrete things (Vol. 3, Ch. 6, Def. 6.23), we have

COROLLARY 4.6. Mental events can cause nonmental events in the same body and conversely.

Consequently disturbances of nonmental (e.g. metabolic) biofunctions may influence mental states and, conversely, mental events, such as acts of will, may influence nonmental bodily states. This is what neurology, neurochemistry, psychiatry, education, and advertising are all about. *Example 1* Yogis learn to regulate their heartbeats, intestinal contractions and even rate of oxygen intake (cf. Miller, 1969). *Example 2* A blow in the head can "knock out the mind" erasing all memory of events immediately preceding the accident. *Example 3* By implanting electrodes in selected sites of the brain, and sending electric currents through them, the neurophysiologist can control some mental states and, through them, elicit or block behavior such as rage (cf. Delgado, 1969).

The most dramatic and revealing of all surgical interventions is of course the sectioning of the bridge between the two cerebral hemispheres, or corpus callosum (Sperry, 1964, 1966, 1977; Gazzaniga, 1967; Bogen, 1969). The split-brain patient behaves in many situations as if he were in two minds – as forecast by Fechner in 1860. For example, while his left hemisphere may wish to read, his right hemisphere may wish to go for a walk, and so a conflict arises within one and the same skull. This is not surprising in our framework: indeed, according to Postulate 4.3 and Corollary 4.4, there is a single mind as long as there is a single plastic supersystem. But if the latter is cut into two then two plastic systems emerge, each with its own mind or system of mental functions:

COROLLARY 4.7. Let b be an animal whose plastic nervous system is split into two detached parts, L and R. Then the mind of b during any time period τ posterior to the splitting divides into two disjoint functional systems:

$$m(b,\tau) = m_L(b,\tau) \cup m_R(b,\tau), \quad \text{with} \quad m_L(b,\tau) \cap m_R(b,\tau) = \varnothing.$$

Hence, *pace* Puccetti (1977), the results of commissurotomy do not support dualism. On the contrary, they show conclusively that no immaterial mind holds the hemispheres together. Dualism could be saved only by invoking divine intervention at the time of splitting the corpus callosum. God would have to replace the original single soul by two new, independent souls – each of which would have to remember something of the past history of the original soul. This looks like an easy solution but it is not for it poses the grave theological problems of which, if any, of the three souls deserves being saved, and what to do on Resurrection Day with three immortal souls and one body. Even assuming that this theological difficulty could be resolved, the dualist is in an awkward position. If he denies that the split brain patient has two minds, he defies experience. And if he acknowledges that a physical entity such as a surgeon's scalpel can slice a mind into two, he contradicts himself (Bunge and Llinás, 1978).

The chemical manipulation of the mind, though less spectacular, is no less clear cut an indictment of dualism. All drugs that act on mental states, from tea to cocaine, do so by modifying the CNS metabolism in some way or other, from changing properties of the neuron membrane to taking part in biochemical reactions. Thus by modifying membrane conduction this drug produces amnesia and that one facilitates recall; and by blocking the synthesis of a certain neurotransmitter this drug produces schizophrenia and that one cures it. (For a list of kinds of deliberate modifications of brain and behavior, see Omenn and Motulsky (1972).)

Finally, the purely behavioral or psychological manipulation of the mind is no less physical, because the sensory inflow and the behavioral output controlled by the teacher (or preacher, propagandist or psychiatrist) modulates neural connectivity, reinforcing certain connections and inhibiting others. In sum, the mental can be controlled, nay is being controlled, in various ways precisely because it is not immaterial: if it were, it would be immune to all attempts at manipulating it.

What holds for the physical control of the mind holds also, *mutatis mutandis*, for the mental control of the body, so often cited as proof of the superiority of Spirit over Matter. Consider three well known phenomena: hunger, anger, and yoga. The process involving the hunger feeling is normally this. Sensors in the digestive tract send signals to the hypothalamus, which in turn alerts the cortex. The motor cortex and the cerebellum control the muscales intervening in the search and reach for food, and in eating. The dualist thinks he can afford to skip all these links: he discards the process ensuing in the cortex's wanting the body to eat, and speaks of the direct action of the mind on the muscles. His account is far too simple – and unscientific. Saying that *X* ate because he *meant* to eat or had the *intention* of eating is just redescribing an ordinary situation in ordinary language terms, not giving an explanation of the process. And yet such accounts are as popular with ordinary language philosophers (e.g. Alston, 1974) as the account of behavior in terms of instincts was among ethologists half a century ago.

Consider anger next. The dualist is satisfied with "explanations" like this one: "She kicked him because she was angered by his words". The physiological psychologist, on the other hand, analyzes this ordinary knowledge account as a causal chain of bodily events: Hearing an utterance → Activation of psychons in the forebrain and of the adrenals → Action of forebrain and adrenals on motor center → Movement of limb. Likewise, when confronted with yoga exercises and biofeedback techniques for moving things without the intervention of any muscles, the physiological psychologist concludes that the autonomic nervous system is not as fully autonomous as we used to think, but can be subjected at least in part to control by cortical systems. There is no evidence for the action of an immaterial mind on the body and there can be no such evidence, for the ghostly is inaccessible to science. Epicurus knew this twenty-two centuries ago.

2.4. *Location of the Mind*

If the mind is immaterial then it makes no sense to ask where it is. The *prima facie* evidence for the nonspatiality of the mind is of this kind: when one thinks about something he does not feel his thoughts to be located anywhere. So much so, that in the past the heart and the liver were assumed to be the "seats" of the mind. However, as long as we remain on the ordinary knowledge level we cannot refrain from citing evidence for the thesis that the mind is coextensive with the body: e.g. when your foot hurts you feel pain there, not in your brain. So, ordinary experience is ambiguous: it supports now the thesis of the nonspatiality of the mental, now the thesis that the mind can wander throughout the body. This suggests that the nonspatiality thesis is confused, and that introspective evidence is insufficient or even irrelevant. We should examine both.

The thesis of the nonspatiality of the mental contains two different hypotheses. One is that mental events are *unextended*, the other that they occur *nowhere* – except of course "in" the mind, which is allegedly immaterial and therefore unextended. The latter hypothesis implies the former but not conversely. Indeed, mental events might be unextended yet occur somewhere – say in the brain. Likewise a physical event that takes no time occurs at some time or other. In our ontology there are no events in themselves but only events in some thing or other (Vol. 3, Ch. 5). Hence the question of the space "occupied" by an event is the problem of the extension of the changing thing. Thus not the *firing* of a neuron but the *firing neuron* is spatially extended.

Events occur wherever the "eventing" (changing) things may be (Vol. 3, Ch. 6, Sec. 4.2). In particular, mental events occur in some plastic neural system or other. So, in principle and often in practice, the neurophysiologist can locate a thought process, e.g. by eliciting it with electrical stimulation (Penfield and Rasmussen, 1950). Hence the ordinary language philosopher's claim that "it makes no sense at all" to talk about mental states and events as being located somewhere in the body has been experimentally refuted. Mental events are just as localized as the corresponding plastic neural systems. If the latter have a fixed location, so have the mental events they engage in. And if the neural system in question happens to be itinerant, the mental event itself will be itinerant – just as much as a light wave. In short, mental events occur in the brain. However, the mind, conceived of as the *set* of all mental events (Definition 4.14), is nowhere.

What about pain in limb, particularly a limb that has been amputated (phantom limb)? The short answer is that, just as we see stars in the sky, not in our brain, so we locate pains in several parts of the body, and sometimes either localization is wrong. A long answer is that the cortex of the adult human being has a map of the entire body (the somatotopic representation). It is here that the signals coming from various parts of the body are decoded and located. If they come from the stump of an amputated limb, the patient can feel them as originating in the missing part because he has formed since childhood a comparatively unalterable map. (There are alternative neurophysiological explanations of phantom experience.)

In sum, mental processes are located wherever the corresponding psychons are located; likewise they occur in time and take some time to happen. This generalization from contemporary neurophysiology – foreshadowed by the Hippocratic school – comes as no surprise for the materialist, for whom the *res cogitans* is a *res extensa*, namely the brain.

What about ideas: are they too in spacetime? The answer to this ambiguous question depends on the construal of 'idea'. If taken as ideation processes, ideas *are* in the brains that happen to think them up – and only there and at the time they are being thought. On the other hand the so called *product* of any such process, i.e. the idea in itself, is nowhere in spacetime because it does not exist by itself: we only feign it does. For example, although thinking of the number 3 is a brain process, hence one located in spacetime, the number 3 is nowhere because it is a fiction existing by convention or fiat, and this pretence does not include the property of spatiotemporality. What holds for the number 3 holds for every other idea – concept, proposition, or theory. In every case we abstract from the neurophysiological properties of the concrete ideation process and come up with a construct that, by convention, has only conceptual or ideal properties. (Cf. Vol. 1, Ch. 1.)

To sum up: Whereas the mental is in some head or other, ideas in themselves, having no being of their own, are nowhere.

2.5. *Mentalist Predicates*

At first blush the most formidable objection to psychophysical monism is this: "The materialist claims that all mental events are brain events, but he cannot offer a clear description, let alone a definition, of a mental event without the help of mentalist concepts such as those of self, privacy, and immediate accessibility. Consequently he cannot even state his identity thesis. In other words the materialist cannot consistently state the identity of phenomenal and physical predicates because he refuses to buy the former at their face value".

Actually the above is not an objection specifically addressed to the psychoneural identity thesis but one that could be raised against all science conceived of as a cognitive enterprise reaching beyond the ordinary language description of appearances. Our rejoinder is this:

(i) Science attempts to account for reality behind appearance, so it either does not employ phenomenal predicates or, if it does, regards them as derived not as basic. Surely phenomena, in particular mental phenomena, are experientially immediate, but they are neither ontologically nor scientifically primary: they are something to be explained.

(ii) Appearances (phenomena) can be explained by psychology, at least in principle, in strictly nonphenomenal terms. For example the various visual illusions can be explained either as results of incomplete or inadequate information, or as results of the fatigue of certain neurons, or the faulty connection of others.

(iii) A rigorous formulation of the monistic thesis must employ no ordinary phenomenal or mentalist predicates – or fuzzy expressions such as 'mental-neural correlation'. (At any rate ours does not.) The identity thesis is that every mental process is a brain process, not that every mentalist sentence is identical with some neurophysiological sentence: the identity is ontological not linguistic. (This important point has been obscured by the nonmaterialist version of the identity theory defended by Schlick (1925) and Feigl (1958).)

In other words, we must not demand that science descend to the level of common sense but, instead, strive to raise the latter to the level of the former. After all, this is what we do when shopping for breakfast cereal and, instead of allowing ourselves to be goaded by the advertisement that "X peps you up", we look at the protein, vitamin, and calorie content of X – or when shopping for a car and, rather than remain satisfied with the assurance that "X is spirited", we inquire into the acceleration of X. We should behave similarly with regard to mentalist predicates. More precisely, we should adopt

RULE 4.1 Whenever possible dispense with mentalist predicates: either replace them by, or define or deduce them with the help of, neurophysiological predicates.

Example of elimination "The idea of an immaterial mind controlling the body is vitalism, no more, no less; it has no place in science" (Hebb, 1974). *Example of reduction* The degree of introversion of a person equals the activity of his or her frontal-medial-septal-area hippocampal system (Gray, 1972a).

Rule 4.1 invites either of the following strategies with regard to mentalist predicates: elimination, definition, or deduction with the help of neurophysiological concepts. Of these the second, i.e. definition (or identification), has been commended by the so-called identity theorists (e.g, Smart, 1959). That is, they have proposed that, for any mentalist predicate M, there is a neurophysiological predicate N such that $M = N$, where $=$ is the relation of ordinary (or strict) identity. Ordinary language philosophers have claimed that this cannot be, for "it makes no sense" (Malcolm, 1964). The obvious rejoinder is that no scientific hypothesis whatsoever makes sense to them, for they insist on remaining within the confines of ordinary language, hence ordinary knowledge.

Another, related objection to the identity thesis, and one commended as definitive by Popper (Popper and Eccles, 1977), is as follows (Kripke, 1971). If the identity is to be taken just as seriously as "Heat is the motion of molecules", then it must be a necessary identity, in the sense that it must hold in all possible worlds – whatever these may be. However, this cannot be, because "it seems clearly possible" (to Kripke) that M (e.g. pain) could exist without the corresponding brain state N, or that the latter could exist without being felt as pain – whence the identity is contingent and therefore flimsy. Rejoinder: (*a*) scientists and science-oriented philosophers do not waste their time speculating about (logically) possible worlds: they want to explore the real world (Bunge, 1977a); (*b*) the difference between necessary (or strict) identity and contingent identity does not occur in ordinary logic or in science. In sum, the sophistic objection to materialism holds no water.

Surely our Rule 4.1 is not a summary of the state of the art but rather a programmatic injunction, i.e. one that can guide research. As such it can be confirmed. And it should be discarded (or simply forgotten) if found barren. But, far from barren, it is behind many a success in contemporary psychology and neuroscience.

Note that Rule 4.1 is the *weak* not the strong reductionist thesis: we favor the *partial* reduction (definition or deduction) of the mental to the neurophysiological, not the elimination of its emergent properties (cf. Bunge, 1977b, 1977c). Thus we do not state that a frog enjoys listening to a Beethoven quartet, or that what a human feels when listening to it is no different from a frog's sighting a fly. We only say that enjoying Beethoven is a process in a human auditory system educated to face such experience.

The partial reducibility (definability or deducibility) of mentalist predicates does not entail that *every* ordinary language sentence concerning mental events is translatable without further ado into a (possibly very complex) scientific formula. Nor, *a fortiori*, does it entail that every mentalist sentence is identical with some neurophysiological sentence. If this were so then there would be a difference only in degree between the scientific and the mythical accounts of the mental. In short, genuine psychoneural monism (unlike the linguistic "identity theory") does *not* assert that, given any mentalist sentence m, there is a neurophysiological sentence n such that $m = n$.

The best one can do with mentalist sentences is either to eliminate them altogether or to try and cleanse, deepen, and refine them. For example, the phrase 'The soul survives', which makes sense in some theologies, makes none in our framework, where *soul* does not occur at all and *mind* is defined as a set of neural processes. (Sets are neither alive nor dead.) And those ordinary language mentalist sentences that are transtatable into neurophysiologese – e.g. 'Her mind is at work', or 'I am having a yellow afterimage' – may be so only roughly. As for the converse translation, of neurophysiological and psychophysiological sentences into ordinary language (mentalist) sentences, it is, more often than not, impossible, as should be obvious from looking at any of the mathematical models of mental processes published over the past few years.

In short, the various languages employed to describe mental events – in particular mentalese, behaviorese and neurophysiologese – are not inter-translatable on the whole. This is because, save exceptions, their sentences do not express the same propositions. For example, 'She is happy', 'She is grinning', and 'Her pleasure center is active' are, though related, quite different. In fact, although they have the same referent (she), and each describes one aspect of one and the same process, they have different senses and therefore are inequivalent.

The inequivalence of most mentalist and neurobiological propositions is no obstacle to the building of bridges between them. The entire science of psychophysics is an attempt to disclose some such bridges, e.g. the relation between sound intensity and felt loudness. (Psychophysical laws are of the form: "For all x, if x is an animal of species K, then: If physical stimulus p impinges on x, then x feels mental phenomenon m". Such generalizations may eventually be deepened by expressing m in neurophysiological terms.) Neurology, too, looks for bridges between the mental (and also behavioral) states reported by neurological patients in phenomenal language (e.g. 'I suffer from migraine'), on the one hand, and sentences describing the corresponding malfunction of the CNS on the other.

(Neurological laws are of the form: "For all x, if x is human, then: If x is in neurological condition n, then x feels mental phenomenon m".)

So much for the general and basic concepts and assumptions of our version of the psychobiological theory of the mental. We proceed now to examining a few particular and rather typical kinds of mental process.

3. SENSATION TO VALUATION

3.1. *Sensing and Perceiving*

All things react to external stimuli, but some react more selectively than others: they are said to detect them. The general concept of specific reaction, or detection, is elucidated by

DEFINITION 4.16 A system *detects* things or events of a certain kind (or is a *detector* of them) if and only if it reacts to them only.

Multicellular organisms have a variety of detectors grouped into systems called *selective systems* (or also, misleadingly, *recognition systems*). For example, the immune system can detect a large variety of molecules because it is composed of detectors (antibodies) of many kinds. And the CNS can detect a huge variety of internal and external events, though by no means all of them.

For an animal it is not enough to possess detectors: the animal must be able to do something about the events it detects. This necessitates amplification and transmission of signals in a uniform fashion, so that each signal can be "read" unambiguously by the CNS. This is what neurosensors or neuroreceptors do. A mammal has myriads of neurosensors: for motion, pressure, heat, rate of secretion, acidity changes, muscle contraction, novelty, etc. This is the general concept:

DEFINITION 4.17 A detector is a *neurosensor* (or *neuroreceptor*) iff it is a neural system or is directly coupled to a neural system.

Example The visual cortices of cat and monkey (and probably also of man) contain neurons specializing in detecting vertical lines and others horizontal lines (Hubel and Wiesel, 1959, 1962). These are called *feature detectors*.

In higher animals neurosensors come in systems, so we need

DEFINITION 4.18 A *sensory system* of an animal is a subsystem of the nervous system of it, composed of neurosensors and of neural systems coupled to these.

The sensory systems of the higher vertebrates are extremely complex systems involving not only sense organs but also portions of the cerebral cortex. Moreover they are closely linked to other subsystems of the CNS, particularly the motor system. And they are not just detectors of environmental inputs, like the thermocouple or the photoelectric cell. Indeed the sensory systems are under the constant action of the CNS by way of the gamma afferent fibers, so that their state depends not only upon external stimulation but also on the state (and history) of the CNS. This explains why we seldom perceive the same stimulus in the same manner. And it shows that (*a*) stimulus-response psychology is basically false, and (*b*) any model of perception contrived in imitation of purely physical or chemical detectors, such as photographic cameras, is bound to fail.

Next comes

DEFINITION 4.19 A *sensation* (or *sensory process*, or *feeling*) is a specific state of activity (or function or process) of a sensory system.

Feeling cold or hot, hungry or sated, tired or energetic, pain or pleasure, and so on, are functions of so many sensory systems. All sensations (or feelings) are then in the brain even though they are usually elicited outside it. Hence sensations can be either blunted or made more acute by manipulating the brain, be it electrically, chemically, or behaviorally. If they were in an immaterial mind we would have no access to them from the outside.

Sensing is not all: in vertebrates sensory information is processed or "interpreted" by special sensory "areas" in the brain cortex. In the primate brain each such sensory cortical "area" can be subdivided into three parts: the primary, the secondary, and the tertiary ones. Only the primary sensory cortical area should be regarded as a component of the corresponding sensory system. Presumably it loses plasticity as the animal develops. On the other hand the secondary and tertiary sensory areas seem to be plastic throughout the animal's life. We shall call them the *plastic* neural systems *directly* attached to the sensory system. We shall assume with Hebb (1968) that, whereas sensation is the specific activity of a sensory system including the primary sensory cortical area, perception is this activity together with the activity elicited by the sensory system in the plastic neural system directly attached to it. See Figure 4.4.

This figure summarizes the following features of human perception:

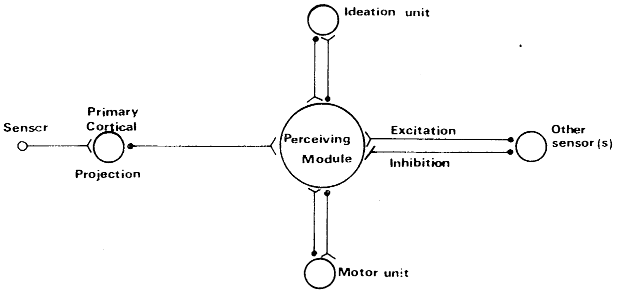


Fig. 4.4. Schema of a perceptual system: a modification of Bindra's (1976), in turn a modification of Hebb's (1949).

(i) the perceiving is done by a neural system located in a secondary sensory projection of the brain cortex;

(ii) the functioning of the central perceiving unit is strongly influenced by some motor unit or other as well as by ideational units and by sensory inflows of various modalities;

(iii) the central perceiving unit can be activated by any other units coupled to it (e.g. dreams, hallucinations, phantom limb experiences, eidetic images);

(iv) perception can guide motion as well as ideation.

We compress the preceding into

DEFINITION 4.20 (i) A *percept* (or *perceptual process*) is a specific function (activity, process) of a sensory system and of the plastic neural systems(s) directly coupled to it;

(ii) a *perceptual system* is a neural system that can undergo perceptual processes.

We assume – in line with Bindra (1976) – that, whereas a sensory system has a roughly constant neuronal composition, the corresponding perceptual system has a variable part. This explains why a given stimulus can activate different perceptions (as in the case of the duck-rabbit figure or of the Necker cube) and why different stimuli can give rise to the same perception (as in the case of size constancy). See Figure 4.5. In other words, two percepts are *equivalent* iff they are elicited by the same sensation and involve the same plastic neural system.

We shall assume that the perception of an external object is the distor-

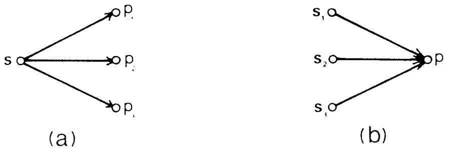


Fig. 4.5. Sensation and perception. (*a*) Same sensation elicits different perceptions according to the plastic neural assembly it activates. (*b*) Different sensations give rise to same perception by activating the same plastic neural system.

tion it causes on the ongoing activity of a perceptual system. To be more precise, we assume

POSTULATE 4.6 Let ν be a perceptual system of an animal b, and call $\pi_s(\nu, \tau) = \{\mathbb{F}^{\nu}(t) \mid t \in \tau\}$ the specific process (or function) that ν engages in during the period τ when in the presence of a thing x external to ν, and call $\pi_s^{\circ}(\nu, \tau)$ the specific function of ν during the same length of time when x fails to act on ν. Then b perceives x as the symmetric difference between the two processes. I.e. the perception of x by b during τ is the process

$$p(x, \tau) = \pi_s(\nu, \tau) \Delta \pi_s^{\circ}(\nu, \tau).$$

This postulate accounts for the active or creative nature of perception and thus for the fact that one and the same perceptual system, when in different states, will perceive differently a given object. This hypothesis disagrees with the direct apprehension doctrine of empiricists and intuitionists (in particular Gestalt theorists). But it is consonant with the current neurophysiology of perception, particularly because it incorporates the spontaneity of the CNS denied by the causal theory of perception. External and internal stimuli do not set the CNS in motion but modulate or control its incessant activity. I.e. the environment enhances or dampens, and in general controls or modulates the activity of the CNS rather than causing it.

3.2. *Mapping Body and Environment*

What do we perceive and how? The naive realist's answer is: We perceive things as they are. Yet imagine Parmenides' homogeneous and eventless universe. An observer would be unable to perceive anything in it, because there would be nothing to discriminate from anything else, and because nothing would be stirring. He would be unable to see, because only things that bounce photons back can be seen; he would hear nothing, because only moving things can generate sound waves; and he would have no tactile sensations, because touching is disturbing – and being disturbed. The observer himself must be changeable, because perception is a process in an animal equipped with perceptual systems.

So, we perceive not things but events, and not all events but only some of those that affect us: the light bouncing off the surface of the lake and hitting our retina, the cry of the owl flying nearby and generating sound waves that set our ear drums in motion, the flame licking our hand, and so on. Whatever we perceive is an event or a sequence of events, and not just any event but one originating in a neurosensor or acting on the latter and, in any case, belonging to our own event space (or the set of events occurring in us). And our perceptions are in turn events in the plastic part of our own sensory cortex. They are not autonomous events but events that map or represent events occurring in other parts of the body or in our environment. Surely this mapping is anything but simple and faithful, but it is a mapping nevertheless. Thus we assume

POSTULATE 4.7 Let b be an animal equipped with a perceptual system c, and call $S(b)$ the state space of b, and $S(c)$ that of c. Moreover, let $E(b) \subset S(b) \times S(b)$ be the animal's event space, and $E(c) \subset S(c) \times S(c)$ its perceptual event space. Then there is a set of injections (one to one and into functions) from the set of bodily events in b to the set of perceptual events in c. Each map, called a *body schema*, depends on the kind of bodily events as well as on the state the animal is in. I.e. the general form of each body map is

$$m: S(b) \times 2^{E(b)} \longrightarrow 2^{E(c)},$$

where 2^X is the family of all subsets of X.

Remark 1 This assumption does not specify the body schemata and does not even state how many there are. Such specifications are tasks for psychobiology. *Remark 2* At first blush phantom limb experiences refute the postulate. They do not, because the body schemata are learned: young children do not have that experience. *Remark 3* The bodily events $x \in 2^{E(b)}$ can be said to be felt by animal b in state s iff b has a body schema m such that $m(s, x) \in 2^{E(c)}$ – i.e. if those events project on to the cortex c of b.

The external world maps are similar to the body schemata. Only, in this case the bodily events are caused by external events, and the maps are far more numerous than the body schemata. In fact there is an entire atlas of the external world, made up of dozens of maps – one for shapes, another for colors, another for motion, and so on. Our assumption is

POSTULATE 4.8 Let $E(e)$ be a set of events in the environment e of an animal b equipped with a perceptual system c, and call $S(b)$ the state space of b and $S(c)$ that of c. Moreover let $E(b) \subset S(b) \times S(b)$ be the animal's event space, and $E(c) \subset S(c) \times S(c)$ its perceptual event space. Then there is a set of partial maps k from sets of external events in $E(e)$ to ordered pairs $\langle$state of b, set of bodily events in $b\rangle$, and another set of partial maps p, from the latter set to sets of perceptual events. Furthermore the two sets of maps are equally numerous, and each map k composes with one map p to form an *external world map* of b in e, or ε. I.e.

$$\varepsilon: 2^{E(e)} \xrightarrow{k} S(b) \times 2^{E(b)} \xrightarrow{p} 2^{E(c)}.$$

DEFINITION 4.21 Let b be an animal with perceptual system c in environment e. Moreover call $S(b)$ the state space of b and $E(e)$ that of e. Then b, when in state $s \in S(b)$, *perceives* external events in $x \in 2^{E(e)}$ if and only if [these cause bodily events that are in turn projected on to the sensory cortex c, i.e. if] $k(x) = \langle s, y\rangle$ with $y \in 2^{E(b)}$ and in turn $p(s, y) \in 2^{E(c)}$. Otherwise the events in x are *imperceptible* to b when in state s [i.e. imperceptible events either do not cause any bodily events or cause them but do not get projected on to the perceptual system].

Remark 1 Just like the body schemata, the external world maps are not point to point but rather set to set correspondences. *Remark 2* There are many external world maps, several for each modality. The various visual maps integrate into one, which constitutes the visual space. Likewise the auditory maps integrate into the auditory space, and so on. *Remark 3* The various sets of external world maps, one for each modality, integrate in turn into the external world atlas. This integration can be explained as the simultaneous activation of the various perceptual systems. Such an activation can be achieved by sensory inflow to one or more receptors, and activation of other perceptual systems via associated neural systems. For example, a total or crossmodal perception of a familiar hand may be had just by its haptic perception, because the tactual receptors activate others. Such integration is learned. *Remark 4* Not only does the organism learn to integrate the activities of its various perceptual systems: it also learns to perceive in each modality. In particular, the young vertebrate learns to touch, to hear, to see, and to smell. Here again the psychobiological viewpoint can teach us something that dualism cannot. In fact from a neurobiological point of view learning to perceive is a process of self-organization of the synaptic junctions in the sensory cortex.

3.3 *Behavior: Definitions and Principles*

The set of motor outputs of an animal, whether global as in locomotion or partial as in grasping or grinning, moving the eyeballs or excreting, is called its behavior. In short, we make

DEFINITION 4.22 For any animal b,

(i) the *behavioral state* of b at time t is the state of motion of b at t;

(ii) the *behavior* of b during the time interval τ is the set of all behavioral states of b throughout τ.

When inherited, or else when learned and successful, behavior can become stable or recurrent. In this case it is often called a behavior pattern:

DEFINITION 4.23 A *behavior pattern* is recurrent behavior.

Animals belonging to different species can do different things, such as swimming of flying, burrowing or building nests, preying or fleeing from predators. And even when they do the same 'thing' (i.e. go through the same motions) they may do so in different fashions or styles. In short there are different kinds of behavior, and every animal species is characterized by some to the exclusion of others. We need then

DEFINITION 4.24 Let b be an animal of species K, and let A be the union of all animal species. Then

(i) the (*possible*) *behavior of type i of animal b*, or $B_i(b)$ for short, is the set of all (possible) behaviors of b associated with the ith biofunction (in particular neural biofunction) of b;

(ii) the *behavioral repertoire* of animal b, or $B(b)$ for short, is the union of all (possible) behavior types of b, i.e.

$$B(b) = \bigcup_{i=1}^{n} B_i(b);$$

(iii) the (*possible*) *behavior of type i of species K*, or $B_i(K)$ for short, is the union of all the (possible) behaviors of the members of K, i.e.

$$B_i(K) = \bigcup_{x \in K} B_i(x);$$

(iv) the *behavioral repertoire of species K*, or $B(K)$, is the union of all (possible) behavior types of K:

$$B(K) = \bigcup_{i=1}^{n} B_i(K);$$

(v) the *specific behavioral repertoire of species K* is the behavioral repertoire exclusive to members of K:

$$B_s(K) = B(K) - \bigcup_{X \subset A} B(X), \quad \text{with} \quad X \neq K;$$

(vi) *animal behavior* is the union of the behavioral repertoires of all animal species, i.e.

$$B = \bigcup_{X \subset A} B(X).$$

(All of the above definitions are understood to refer to a given stage in the development of the individual or the evolution of the species. As an animal develops its behavioral repertoire changes. Something similar holds for the evolution of behavior.)

Our first assumption is that behavior, far from being primary, is derivative:

POSTULATE 4.9 The behavior of every animal endowed with a nervous system is controlled ("mediated", "subserved") by the latter. I.e. for every behavior type B_i of animals endowed with a nervous system, the latter contains a neural subsystem that controls the motions in B_i.

An immediate consequence is

COROLLARY 4.8 Any change in (nonredundant) neural systems is followed by some behavioral changes.

Now, there are no two identical nervous systems – nor, for that matter, two physical systems identical in all respects. This generalization, together with Postulate 4.9, entails

THEOREM 4.1. No two animals behave in exactly the same way [even if they are conspecific].

The assumption that a nervous system may contain a plastic subsystem (Postulate 4.1), together with Postulate 4.9 and Definition 4.24, entails

THEOREM 4.2. The behavioral repertoire of an animal endowed with plastic neural systems splits into two parts: the one controlled by the committed (or prewired) part of the NS of the animal, and its complement, i.e. the behaviors controlled by the plastic components of the NS.

DEFINITION 4.25. The part of the behavioral repertoire of an animal, that is controlled by the committed part of its NS, is called its *inherited* (or *instinctive, stereotyped, modal* or *rigid*) repertoire, while the one controlled by the plastic part of its NS, its *learned* repertoire.

Remark 1 'Inherited' is not the same as 'inborn'. Only the adult animal is in possession of its full inherited behavioral repertoire. (That is, *Inborn* $\subset$ *Inherited* $\subset$ *Total.*) This distinction was not adopted by Chomsky (1968) when claiming that universal grammar is an "innate schematism".
Remark 2 Because natural selection eliminates most types of behavior resulting in backfiring, instinctive or programmed behavior looks marvelously suited to goals, i.e. purposeful. But, as Spinoza (1677, First Part, Appendix) saw, this is sheer anthropomorphism, and contemporary biologists and psychologists are finding less and less use for final causes in explaining behavior.

An immediate consequence of Theorem 4.2 and Definition 4.25 is

COROLLARY 4.9. The behavior of an animal deprived of plastic neural systems is totally stereotyped.

Most inherited behavior has been selected because it is adapted to a certain environment, i.e. it is biovaluable in it. (Not all: recall the moth's compulsion to circle about a flame.) But if the environment were to change radically, some of the behavior types, formerly adaptive, would be devalued. Joining both statements we have another axiom:

POSTULATE 4.10 Provided the environment does not change radically during the lifetime of an animal, most of its inherited behavioral repertoire has a positive biovalue for it.

The great biovalue of possessing a nervous system with plastic neural subsystems is, of course, its capacity to learn new behavior types and thus improve its chances of surviving environmental changes. So, we make

POSTULATE 4.11 Some of the inherited capabilities of an animal endowed with plastic neural systems are modifiable by learning.

So much for our behavioral principles. Now for drives. Explanations of the type "X eats (or sleeps, or mates) because X has an urge (or drive) to eat (or sleep, or mate)" are not illuminating as long as the drive concept remains an occult hypothetical construct. They become genuine explanations the moment drives are construed in physiological terms, e.g. as detection of imbalances (deficits or superavits) of certain physiological variables such as sugar or noradrenalin levels in the blood. In this case the hypothesis that behavior X is caused by drive Y can be put to the test, namely by manipulating the imbalance and watching behavior changes. The entire drive theory can be compressed into one definition and one assumption:

DEFINITION 4.26. A *drive* (or *motivation*) of kind X is detection of an imbalance in the X component(s) of the state function of the animal. More precisely: the intensity $D_X(b, t)$ of drive X in animal b at time t equals the absolute value of the difference between the detected and the normal values of X for b at t.

POSTULATE 4.12. For every drive in an animal there is a type of behavior of the animal that reduces that drive (i.e. that decreases the imbalance in the corresponding property).

The adoption of this hypothesis allows one to explain many behavior patterns without resorting to teleology. Thus a bird engaged in nest building exhibits inner-driven behavior (and, more precisely, a behavior driven by neuroendocrine imbalances), The animal does not allow random sensory stimuli to divert it from its job. (For other examples of the physiological approach to motivation, see Gross and Zeigler (Eds.) (1969).)

In Ch. 3 we stated that certain states and events are objectively valuable to the organism whereas others are not. We now go a step further and affirm that all animals are equipped with receptors enabling them to evaluate some incoming stimuli as nocent, favorable, or indifferent. This does not entail that all evaluations are correct, let alone that all animals are conscious of such evaluations and can make value judgments. Only a few higher mammals can form value judgments: in all others, and even in man most of the time, valuations are automatic. What distinguishes man from other animals, with regard to valuation, is that he can reason about values as well as evaluate reasons. Other animals do not have the proper CNS for performing such syntheses of reason and value: they behave without weighing behavior and, if they reason at all, they do not weigh reasonings.

In any case, saying that an animal is capable of evaluating items of a certain kind amounts to saying that it has a value system – a concept characterized by

DEFINITION 4.27. Let S be a set of items and b an animal. Further, let $\succsim_b$ be a partial order on S. Then the structure $\mathscr{V}_b = \langle S, \succsim_b \rangle$ is a *value system* for b at a given time iff

(i) b can detect any member of S and discriminate it from all other items in S;

(ii) for any two members x and y of S, b either prefers x to $y (x \succsim_b y)$ or conversely $(y \succsim_b x)$ or both $(x \sim_b y)$ at the given time.

Remark 1 This is a comparative concept of value. A quantitative concept is that elucidated by utility theory and occurring in decision theory. *Remark 2* Note the difference between the concept of psychovalue elucidated by Definition 4.27 and that of biovalue introduced in Ch. 3, Sec.2.2. The biovalue of a for b at t is the objective value that a *possesses* for b at t. On the other hand the psychovalue of a for b at t is the value that b *assigns* to a at t. These value assignments may be biologically mistaken, in which case psychovalues conflict with biovalues. The psychophysical dualist would regard this as a conflict between mind and body, while the monist would regard it as a conflict between the brain or a subsystem of it, on the one hand, and the rest of the body or a different brain subsystem, on the other. In any case, since such conflicts do occur, we had better not try to reduce one kind of value to the other even though both valuations are biological processes. *Remark 3* Value systems are known to change for one and the same organism, not only in the course of its development but also with circumstances. Therefore we incorporate this fact into our hypothesis:

POSTULATE 4.13. All animals are equipped with a value system, and those capable of learning can modify their value systems.

Finally we introduce the notion of choice. An animal presented with a set of alternatives may abstain from choosing among them or may pick one of them. If it tries out some of the options and finds out the hard way which lead to the most valuable outcomes, the animal will be said to have learned to choose. But in all cases there must be some freedom in order for there to be genuine choice. I.e. the animal must be able to choose any of the alternatives open to it, and some choices must be (at least initially) mistaken (i.e. disvaluable) for there to be choice learning. Hence we make

DEFINITION 4.28 Let $\mathscr{V}_b = \langle S, \succsim_b \rangle$ be a value system for an animal b at a given time, and call $A \subset S$ a set of alternatives open to b, i.e. belonging to the behavioral repertoire of b at the time. Then b *chooses* (or *selects*) option $x \in A$ iff

(i) it is possible for b to pick (i.e. to do) any alternative in A(i.e. b is free to choose);

(ii) b prefers x to any other options in A; and

(iii) b actually picks (i.e. does) x.

Note the difference between preference and choice: the former underlies and motivates the latter. Choice is valuation in action, or overt valuation – hence an indicator of valuation not a definition of the latter. And note also that not every choice implements a decision. Decisions are deliberate or reasoned (even if not always rational), and reasoning is the privilege of only a few animal species. Most choices, even in daily human life, are not preceded by any decision-making process.

4. RECALL TO KNOWLEDGE

4.1. *Memory and Learning*

Many a complex prebiotic system, such as a sword and a magnet, has memory of some past states. If only for this reason we must treat memory separately from, and even prior to, learning. A general definition is

DEFINITION 4.29 A system σ at time t has *memory of* (or *memorizes*) some of its past states iff the state of σ at t is a function(al) of those past states.

A system with memory can be called a *memory system*. A memory system records ("commits to memory"), stores, and recalls when suitably stimulated. (Actually the second property may be nonexisting: instead of keeping a record, trace or engram of a state, a system may acquire just the disposition or propensity of recalling the state.) All organisms are memory systems, if only because they have genes, and these record part of the history of the species. Besides genetic memory, all organisms have at least some sort of short span memory: otherwise they would not live long enough to reproduce. So, we can assume

POSTULATE 4.14 All animals have memory of some of their past states, and none of all of them.

There is probably a large variety of memory systems or mechanisms.

We may class them into fixed and itinerant. While invertebrates and lower vertebrates have probably only fixed memory systems (i.e. systems with constant neuronal composition), higher vertebrates have also itinerant memory systems, i.e. systems that are formed just for the occasion. The latter are particularly interesting in the case of man, where remembering seems to consist not in the reactivation of some fixed neural circuit but "in the fresh production or reconstruction of an item" (Bindra, 1976, p. 330). In any case, memorizing is not passive recording but a kind of activity of certain neural systems possibly involving the strengthening of synaptic connectivities (Anderson, 1972, 1973). Nor do we "store" past events in our memory, let alone in our mind: if we store anything at all it is in our brain.

Some memory systems have the ability to learn. According to Definition 4.9, learning is the acquisition of new neural functions, whence it involves plastic (not only wired-in) neural systems. This definition differs from the usual behavioral definition of learning as modification of behavior in response to stimulation. The behaviorist definition is concerned with the effects rather than the causes, and it could be applied to any thing, for any thing will behave differently if stimulated differently.

To emphasize the idea that learning is a modification of neural activity, we make

DEFINITION 4.30 Call E a kind of event or process in a neural system of an animal a involving a plastic subsystem, and S a kind of stimuli (external or internal) which a can detect. Then a has *learned* $e \in E$ in the presence of $s \in S$ during the time interval $[t_1, t_2]$ iff

(i) e did not occur in a in the presence of s before t_1;

(ii) after t_2, e occurs in a whenever a senses s [i.e. a has memorized e].

Since all behavior is controlled by some neural system (Postulate 4.9), the previous definition embraces the concept of behavioral learning, i.e. learning to perform a movement.

We shall also adopt

DEFINITION 4.31 The *experience* of an animal at a given time is the set of all it has learned up until that time.

That is, the experience of an animal is its accumulated learning, i.e. the collection of changes in its plastic neural systems. These include of course nonperceptual learning. On the other hand "learning" to avoid harmful stimuli is not learning proper according to our definition, for it may well consist in a process involving no plastic neural systems – as is indeed the case with invertebrates. The same holds for habituation, which consists in the weakening of existing (wired-in) synaptic connections with repeated stimulation.

Learning, even of the simplest task, is creative for it consists in the appearance of new neural connections and therefore new neural functions. An animal in the process of learning expands its neural state space. This expansion can be either quantitative (i.e. keeping the same axes) or qualitative (changing some axes). A baby's forming its first sentence is presumably an example of the latter type of expansion, whereas its adding one more word to its verbal repertoire is an instance of the former. In general, discovering and inventing are of the second kind, and only a few higher vertebrates seem capable of it.

4.2. *Expectation and Purpose*

So far we have dealt only with non-anticipatory systems, i.e. systems whose behavior does not depend on the expected outcome of their actions. We now introduce anticipation, a property only animals of some species have, namely those capable of expecting either reward or punishment when presented with certain stimuli or engaging in certain activities. We first adopt

DEFINITION 4.32 Animal *b* *expects* (or *foresees*) a future event of kind *E* when sensing an (external or internal) stimulus *s* while in state *t*, iff *b* has learned to pair *s* and *t* with an event of kind *E*.

Because foresight depends upon learning, the lower animals, which have no learning ability, have no expectations either. But if an animal can learn, then it does have expectations, and in turn expecting helps (motivates) learning. Animals capable of expecting can regulate the effort they put in doing something.

It seems that all vertebrates have organs serving an anticipatory or preparatory function. Examples: the salivary glands secrete saliva at the sight of food, and the glucoreceptor cells in the intestine signal to the pancreas the intake of glucose, which stimulates it to release insulin. In both cases one subsystem alerts another, and the anticipatory functions are autonomous or nearly so. No wonder that the mammalian brain too has anticipatory functions and more refined ones: they enable some animals to foresee and to prepare themselves for what may come. These animals are said to behave in a goal-seeking or purposive way. Because expectation presupposes learning, only learned behavior can be purposive and, being learned, it is also motivated: the outcome of the purposive action is expected to bring a reduction in drive. In sum, we make

DEFINITION 4.33 An action X of an animal b has the *purpose* or goal Y iff

(i) b may choose not to do X;

(ii) b has learned that doing X brings about, or increases the chances of attaining, Y;

(iii) b expects the possible occurrence of Y upon doing X;

(iv) b values Y.

Notice the conditions for purposiveness: freedom, learning, expectation, and valuation. Obviously machines, no matter how sophisticated, do not fulfill all four conditions, hence cannot be goal-seeking – except of course by proxy. Most animal behavior is non-purposive: it only looks goal-seeking because it is so often efficient, but this efficiency is the result of control mechanisms that work without the intervention of any plastic neural systems, hence without learning or expectation.

The claim that purposes defy scientific explanation and call for the hypothesis of an immaterial mind was downed by cyberneticians, who suggested a precise general mechanism of purposive action, namely the negative feedback loop (Rosenblueth *et al.* 1943, 1950). Since then neural modellers and physiological psychologists have been able to explain some specific purposive behavior patterns. According to such models purposes are not states or entities in the mind but certain patterns of neural activity. This view, rather than the mentalistic conception of purpose, agrees with evolutionary biology, which is concerned with explaining the emergence of goal-seeking behavior as a latecomer that has nothing to do with immaterial entities.

Having defined the notion of a goal we can now define that of a means:

DEFINITION 4.34 An action X of an animal b is a *suitable means* for attaining a goal Y of b iff in fact b's performing X brings about, or increases the probability of occurrence of, Y.

4.3. *Thinking*

Forming concepts, propositions, problems, and directions are examples of thinking; so are comparing and transforming concepts, propositions, problems, and directions. We assume that thinking of any kind is an activity of some plastic neural system and, as such, one sensitive to minute chemical changes and changes in basic properties of neurons, such as membrane permeability. For example, humans unable to oxydize the amino acid phenylalanine cannot think at all, and normal subjects cannot think straight when in states of extreme stress – which are often states of hormonal imbalance – or when under the action of psychotropic drugs. Believers in the immateriality of reason must ignore the mounting evidence for the hypothesis that the content and level of performance of thinking depend upon such chemical and physiological variables – and, of course, to be consistent they must never resort to such stimulants as caffeine, nicotine and ethyl alcohol.

We shall deal here only with two basic thought processes, namely concept attainment and proposition formation. We shall conceive of the former as the process of forming kinds, such as the class of cats or that of triangles. And we shall conjecture that forming a concept of the "concrete" kind – i.e. a class of real things or events – consists in responding uniformly to any and only members of the given class. Thus we adopt

POSTULATE 4.15 Let C be a set of (simultaneous or successive) things or events. There are animals equipped with psychons whose activity is caused or triggered, directly or indirectly, by any member of C, and is independent of what particular member activates them.

DEFINITION 4.35 Let C be a class of things or events, and b an animal satisfying Postulate 4.15, i.e. possessing a psychon that can be activated uniformly by any and only a member of C. Then *b attains a concept $\theta_b(C)$* of C (or *conceives* C, or *thinks up* C) iff the activity (process, function) stimulated by a C in that psychon of b equals $\theta_b(C)$.

Consider next the formation of propositions. We conjecture that the operation involved is that of psychon pairing. For example, "Children are helpless" pairs "Children" off to "are helpless". Two psychons, possibly cortical columns, are activated sequentially, one for each concept. The sequential activity of the two component system constitutes thinking the corresponding proposition. The idea is encapsulated in

POSTULATE 4.16 Thinking up a proposition is (identical with) the sequential activation of the psychons whose activities are the concepts occurring in the proposition in the given order.

POSTULATE 4.17 A sequence of thoughts about propositions is (identical with) the sequential activation of the psychons whose activities are the propositions in the sequence,

This holds for complex propositions (such as conjunctions) as well as for "trains of thought", in particular inferences. *A fortiori*, it holds for rational thinking, or thought controlled by certain master thoughts, or rules, such as "Abhor contradiction", "Check what follows", and "Supply evidence or counterexample".

In line with our semantics (Vols. 1 and 2) we have distinguished constructs (such as concepts and propositions) from thoughts about them as well as from linguistic expressions. Unlike constructs, thoughts are brain processes. Therefore no two thoughts can be exactly identical. Nobody thinks twice of the number 5, or of the Moon, in exactly the same manner, if only because we never go through exactly the same states. However, we can assume that all the processes of thinking of the number 5 (or of any other construct) fit the same neural pattern, i.e. are equivalent in an essential respect. This notion is elucidated by

DEFINTION 4.36 Two thoughts are *equivalent* iff they consist in thinking of the same constructs. I.e.

$$\theta_a(C) \sim \theta_b(C') \quad \text{iff} \quad C = C' \text{ for any animals } a \text{ and } b.$$

It would have been desirable to define a construct as a set of equivalent neural events instead of presupposing the concept of a construct, for it would root constructs on neural events while preserving their abstract quality. Unfortunately we have been unable to find the proper equivalence relation that would allow one to construct such a definition.

4.4. *Cognition and Decision*

All cognition is of some object, concrete or conceptual, and it consists in some information about its object – complete or partial, true or false. Cognition can be behavioral (e.g. knowing how to ski), perceptual (e.g. knowing the song of the lark), or conceptual (e.g knowing that the lark song is related to its reproductive cycle or its territorial defense). If behavioral, cognition is a disposition of the motor control system to steer motions of a certain kind; if perceptual it is a potential net of percepts, and if conceptual a potential system of concepts. In either case cognition is a brain process. We summarize the preceding in

DEFINTION 4.37 Let a be an animal. Then

(i) if b is a learned behavior type (or pattern), *a knows how to do* (or perform) b iff b is in the actual behavioral repertoire of a;

(ii) if c is a construct, then *a knows c* iff a thinks up (or conceives) c;

(iii) if e is an event, then *a has knowledge of e* iff a feels or perceives e, or thinks of e.

We can use the concept of knowledge to elucidate that of decision:

DEFINITION 4.38 Let x be an arbitrary member of a set A of alternatives accessible to an animal b with the value system $\mathscr{V}_b = \langle S, \succsim_b \rangle$. Then b *decides to choose x* iff

(i) b has knowledge of every member of A;

(ii) $A \subseteq S$ (i.e. b prefers some members of A to others);

(iii) b in fact chooses x.

A possible decision mechanism is this. If the animal perceives or thinks of a number of objects, and is motivated to choose among them, then he will first compare them. Presumably each object will be apprehended by one psychon at a time, and the excitation levels of the various psychons will differ. If one of them is greater than the others, that one will prevail and will send the strongest (or the only) signal to a volition psychon.

The ability to make decisions is restricted to animals capable of knowing. But not all knowledge is of the same grade, nor all valuation correct. When they are, they constitute the basis of rational decision:

DEFINITION 4.39 A decision made by an animal is *rational* iff it is preceded by

(i) adequate knowledge and correct valuations, and

(ii) foresight of the possible outcomes of the corresponding action.

DEFINITION 4.40 A *rational animal* is one capable of making some rational decisions.

It is quite possible that rational decisions are not the exclusive property of man: some subhuman primates seem capable of making rational decisions. And it is sure that no animal is rational all the time.

4.5. *Creativity*

All higher vertebrates are creative, and man is so superlatively. This may be conceded while at the same time it is doubted whether the notion of creativity is clear. The notion is clear if construed as *neural novelty*, i.e. emergence of new neural systems or new functions of existing neural systems – due e.g. to new connections. Such neural novelties may or may not have behavioral manifestations. If they do one speaks of *adaptive behavior*, in particular imitation and invention, according as the novelty is in the particular animal or in the entire species; otherwise one speaks of mental creativity. So, we make

DEFINITION 4.41 Let a be an animal of species K with behavioral repertoire $B(K)$ at time t. Then

(i) *a invents* behavior type (or pattern) b at time $t' > t$ iff a does b for the first time, and b did not belong to $B(K)$ until t';

(ii) *a invents* construct c at time $t' > t$ iff a knows c for the first time at time t', and no other animal of the same species knew c before t';

(iii) *a discovers* event e at time $t' > t$ iff a has knowledge of e for the first time at time t', and no other animal of the same species had such knowledge before t';

(iv) a is *creative* iff a invents a behavior type or a construct, or discovers an event before any other member of its species;

(v) a is *absolutely creative* iff a creates something before any other animal of any species.

Now the hypothesis:

POSTULATE 4.18 Every creative act is the activity, or an effect of the activity, of newly formed neural systems.

Although creativity is maximal in humans, it is likely to be found among all mammals and birds – if only the theological and empiricist prejudice against the concept of creativity is overcome and the phenomenon is investigated scientifically, particularly among young animals placed in unusual circumstances. We make bold and assume that all higher vertebrates are creative:

POSTULATE 4.19 All animals endowed with plastic neural systems are creative.

However, so far as we know only humans are absolutely creative, i.e. capable of inventing or discovering certain "things" before any other animals. We shall use this assumption in defining humanness in Sec. 5.

What is the ontological status of the so-called creations of the human mind? To obtain a sensible answer we must first distinguish the creative process – whether in science, technology, art, or any other field – from its public materializations, such as books, artworks, and machines. Every creative process is a process in some brain or other (Postulate 4.18). In a few cases such mental processes are externalized as physical things or processes which, when perceived by competent observers, elicit in their own brains processes similar to those undergone by the creators. Thus a musical piece can be whistled, performed on a musical instrument, listened to, written down or read. Novels and theories, artifacts and constitutions are similar. What becomes of a musical piece between performances or listening sessions, or after everybody has forgotten it? It does not exist. If the memories of the piece have faded in all brains, and if all scores and discs have been destroyed, the music itself is no more and has no chance of resurrection – except for the most unlikely chance of being reinvented. But if some traces of it do remain – either in brains or on paper, discs or tapes – then the musical piece has the possibility of "coming alive" again. Something that cannot be played on an instrument, an audio system or something similar, and cannot even be hummed, is not a musical piece.

I submit that the same holds, *mutatis mutandis*, for all cultural objects. Thus a sculpture that nobody looks at is just a chunk of matter – and so is a philosophical treatise that nobody reads. There is no immortality in cultural creations just because they can be externalized ("embodied") and catalogued. Only that endures which is being recreated – reperceived, refelt, rethought, re-enacted. Deserted libraries, museums, art galleries, and laboratories, are just cultural cemeteries. Nothing is more dependent and vulnerable – less autonomous – than the "world" of culture. (Let the current contempt for basic science prevail, and soon there will be no scientific culture left.) Culture lives neither in cultural artifacts, such as books, nor in an immaterial and autonomous world of ghosts: it lives in the brains of those who care about it – who cultivate it.

Ideas, then, do not exist by themselves any more than pleasures and pains, memories and flashes of insight. All these are brain processes. However, nothing prevents us from *feigning* that there are ideas, that they are "there" up for grabs – which is what we do when saying that someone "discovered" such and such an idea. We pretend that there are infinitely many integers even though we can think of only finitely many of them – and this because we assign the set of all integers definite properties, such as that of being included in the set of rational numbers. Likewise we make believe that every deductive system contains infinitely many theorems – and this because, if pressed, we can prove any of them. All such fictions are mental creations and, far from being idle or merely entertaining, are an essential part of modern culture. But it would be sheer animism to endow such fictions with real autonomous existence and causal efficacy – the way idealist philosophers from Plato on have been doing, or the way Popper (1968, 1972, 1974, 1977) does. A differential equation does not exist by itself and therefore is harmless; only thinking of it can have some effects. In short, ideas in themselves are fictions and as such have no physical existence.

5. SELF TO SOCIETY

5.1. *Awareness, Consciousness, Will*

While many mentalists equate consciousness with mind, and behaviorists have no use for either concept, physiological psychologists, neurologists and even ethologists are feeling the need for both. For example, the mere description of yoga experiences, such as the voluntary control of heartbeats, uses the concept of initial conscious control. Unfortunately the concept of consciousness is still rather vague – particularly among those who advocate its use but refuse squarely to define it (e.g. Popper and Eccles, 1977).

We start by distinguishing consciousness from awareness. We say that an animal is (or fails to be) *aware* of what goes on in its environment or in itself (in particular, of what it does or is done to it). Only if the animal is aware of some of its own brain processes (not necessarily mental ones) shall we say that it is *conscious*. More precisely, we make

DEFINITION 4.42 If b is an animal,

(i) b is *aware* of (or *notices*) stimulus x (internal or external) iff b feels or perceives x – otherwise b is *unaware* of x;

(ii) b is *conscious* of brain process x in b iff b thinks of x – otherwise b is unconscious of x.

Whereas awareness requires only a sensory system, consciousness requires also intelligence, i.e. the ability to think, and moreover to think of brain events. The lower animals can be aware but not conscious, and a conscious subject may be unaware of some external stimuli. So, we make

POSTULATE 4.20 All animals are aware of some stimuli and some are also conscious of some of their own brain processes.

DEFINITION 4.43 The *consciousness* of an animal *b* is the set of all the states of the CNS of *b* in which *b* is conscious of some CNS process or other in *b*.

In other words, consciousness of brain event *x* is direct knowledge of *x*. (This is what introspection is all about.) Moreover we may hypothesize that a conscious event is a brain activity consisting in monitoring (recording, analyzing, controlling, or keeping track of) some other brain activity, much as a voltmeter measures the electromotive force between two points of an electrical network. I.e. we adopt

POSTULATE 4.21 Let *P* be a subsystem of the CNS of an animal *b* engaged in a mental process *p*. Then the CNS of *b* contains a neural system *Q*, other than *P* and connected with *P*, whose activity *q* equals *b*'s being conscious (thinking) of *p*.

This hypothesis is far from outlandish: after all, every biosystem is a control biosystem, i.e. it monitors (watches and checks) its own activity. By sensing and correcting its own functioning, the biosystem assures normality or return to it. So, it should come as no surprise that the brain, the most complex of biosystems and the supreme control of the animal, is capable of controlling its own activity.

Since by hypothesis two distinct and interconnected neural systems are involved in all conscious states, it is perfectly possible for them to interact. Thus if a subject is conscious of a routine mental task, he or she may bungle it: the monitoring interferes with the execution of the well learned task. If on the other hand the subject is beginning to learn a task, he had better be both aware and conscious of it: the monitoring system will guide the establishing of the psychon. To account for such phenomena there is no need to invoke an immaterial consciousness (as Sperry, 1969, Popper and Eccles, 1977, and so many others have done). If one wishes to speculate on the causal efficacy of conscious events, then he ought to do it within the scientific framework of changing things instead of the mythical framework of disembodies entities.

Since awareness and consciousness are activities of a biosystem, they can be dimmed or heightened by appropriately inhibiting or stimulating the neural components involved. Such degrees are nothing but the intensities of the activities of the corresponding neural control systems. If consciousness were a separate entity we would not know how to explain such variations and, in particular, the daily obliteration and resurgence of consciousness.

It is a well known law that, as learned tasks are repeated, they tend to become automated, i.e. unconscious:

POSTULATE 4.22 In the course of the life of an animal capable of learning, learned behavior, if initially conscious, becomes gradually unconscious.

When automatisms prove insufficient to solve a problem or execute a task, it proves convenient to become conscious or to heighten the level of conscious mental activity. A conscious animal is better equipped to cope with novelty, for it can evaluate its own thoughts and actions, criticizing and correcting them if necessary. (For the adaptive advantages of consciousness see Griffin, 1976.) Given the biovalue of consciousness as a device for instant adaptation, it is no wonder that man has been so successful, for his CNS is the only one capable of generating conscious states almost at will.

Consciousness, then, is not an entity but a set of states of a highly evolved CNS. Therefore to speak of 'states of consciousness' is sheer reification: there are only conscious (and unconscious) states of the brain. (Recall Fig. 4.3 in Sec. 2.2.) Likewise it is mistaken to speak of the Unconscious (or the Subconscious) as an entity, in particular one capable of influencing Consciousness (another entity). There are only brain events, some conscious and the others unconscious, and because they are concrete events they can influence others. What holds for consciousness and unconsciousness holds also for Freud's Id and Superego. There can be no mental entities within mental entities, because mental entities are nonentities. Surely there are controls, e.g. feedback systems, in the brain. Some such controls may "censor" (inhibit) certain brain activities, e.g. block the flow of information to the system doing the monitoring, or the activity of the psychon(s) doing the thinking. For example one may have learned an item A and, in the course of mental work, think an item B incompatible with A. If A is deeply ingrained, it may "repress" or "suppress" B, i.e. inhibit the psychon that does B. But if B is as vivid as A, further psychons may be activated until the "contradiction is resolved" in favor of one of the original options or of a third one. Physiologically, this is nothing but an interplay among competing neural feedback circuits.

Our next topic is volition, or conscious purposeful behavior:

DEFINITION 4.44 An animal act is *voluntary* (or *intentional*) iff it is a conscious purposeful act – otherwise it is *involuntary*.

The will is not an entity but a neural activity: x wills y iff x forms consciously the purpose of doing *y*. Nor is the will a mysterious faculty of an immaterial mind, but a capacity of highly evolved CNSs, namely "a control of behavior by the thought process" (Hebb, 1968, p. 75). When you pick a fruit from a tree or a book from a shelf, and are aware of what you are doing and moreover have a definite purpose in doing so, what happens is that certain psychons in your associative cortex – most likely psychons located in the frontal lobes – activate certain motor centers (in particular the gyrus), which in turn control the movements of the limbs. Thus *intentions can cause behavior*. Curiously enough this is also a thesis of the dualist philosophers who defend the "intentionalist" (as opposed to the scientific) explanation of human behavior. However, they have no right to do so because causal relations obtain only between events, and the dualists refuse to regard intentions as (neural) events.

Voluntary acts can be free or compelled. The general who decides to launch an attack may act freely, but those of his soldiers who go unwillingly to battle act voluntarily though under compulsion. Free will is volition with free choice of goal, with or without foresight of possible outcome. We make then

DEFINITION 4.45 An animal acts of its own *free will* iff

(i) its action is voluntary and

(ii) it has free choice of its goal(s) – i.e. is under no programmed or external compulsion to attain the chosen goal.

Free will has of course been either affirmed or denied for theological or philosophical reasons. In our theory it follows from Postulate 4.20 and Definitions 4.44 and 4.45:

THEOREM 4.3 All animals capable of being in conscious states are able to perform free voluntary acts.

If consciousness is not exclusively human, neither is free will. And both are legitimate subjects of scientific research.

5.2. *Person and Self*

There are a number of concepts of a person. We are interested in this one:

DEFINITION 4.46 If *b* is an animal endowed with a plastic neural system capable of mentation (i.e. with a nonvoid mind), then

(i) the *personality* of b is the functional system composed of all the motor and mental functions of b;

(ii) a *person* is an animal endowed with a personality.

Remark 1 It follows that not only humans but all animals with mental abilities are entitled to be called 'persons'. *Remark 2* Personality is a whole-body property: there is no special organ of personality, since it is defined as the union of behavior and mentation. *Remark 3* For the same reason the destruction or ablation of considerable portions of the CNS, as well as the amputation or the paralysis of limbs, ensue in personality changes. Irreversible and massive destruction of the brain destroys personality altogether, in which case the animal ceases to be a person both psychologically and legally. *Remark 4* Learning shapes and often enriches personality; it can even produce radical personality changes. *Remark 5* People who undergo shattering experiences (war, concentration camp, prison, or ideological conversion) may acquire whole new personalities. *Remark 6* In deep sleep or in coma we lose much if not all of our personality: we become nonpersons or nearly so. (But of course when reawakening we recover – or rather reconstruct – our personality.) *Remark 7* Since the behavior and the mentation of animals endowed with plastic neural systems depend partly on their environment, one and the same animal is bound to exhibit different personalities in different environments – e.g. a tyrant at home and meek at work. (For the various determinants of personality see Gray, (1972b).) *Remark 8* There is no personal identity or enduring personality any more than there is digestive or cardiovascular identity: the behavior and the mentation of a higher vertebrate are more changeable and vulnerable than any other bodily function. (For an early criticism of the idea of personal identity see Hume, 1739, Part IV, Sec. VI.) *Remark 9* Since mind is not incorporeal stuff but a collection of brain functions, and since there are no two strictly identical brains, "identical" twins reared in the same family have different minds, hence are different persons instead of constituting (as Shaffer, 1977, has claimed) a single mind with two bodies. *Remark 10* Since split brain patients have two minds (Corollary 4.7), they also have a double personality, i.e. they are two different persons united anatomically. *Remark 11* If an individual were to get a new brain by transplant, the personalities of both donor and recipient would disappear, and a third personality would emerge. Indeed, the brain in its new encasement would receive new stimuli and fail to receive some of the customary stimuli. And, because it would be controlling a different body, it would effect different movements, so it would function differently. The original selves would be lost. The same holds, *a fortiori*, for the transplant of each hemisphere into a different skull: two different persons would be manufactured – possibly neither of them viable. In short, tampering with a minding body involves tampering with its mind.

Finally we come to the notion of self, which in our theory is elucidated by

DEFINITION 4.47 An animal

(i) has (or is in a state of) *self-awareness* iff it is aware of itself (i.e. of events occurring in itself) as different from all other entities;

(ii) has (or is in a state of) *self-consciousness* iff it is conscious of some of its own past conscious states;

(iii) has a *self* at a given time iff it is self-aware or self-conscious at that time.

Remark 1 This concept is more restrictive than that of mind: all animals possessing a self (i.e. all selves) mentate but the converse is false. *Remark 2* The self is not an entity but a state of an entity, namely an advanced brain. Therefore in the context of our theory to say that the self has or uses its brain (Toulmin, 1972; Popper and Eccles, 1977) amounts to saying that certain brain states have a brain. *Remark 3* Possibly animals other than humans have self-awareness but, so far as we know, only humans have *self*-consciousness, at least when normal and past their infancy. The young child is self-aware but not self-conscious. Self-consciousness is generally believed to appear at about seven years of age (Piaget) and is sometimes conjectured to originate as internalized speech, hence as a spinoff of social life (Vygotskii, 1962; Luria, 1969). It is therefore as old as language, and both have evolved (and continue to evolve) together with society. But the subject of sociality deserves a new subsection

5.3. *Social Behavior*

A bird's territorial call and a criminal's "antisocial" action are no less social than a bee's dance or a baboon's scouting on behalf of its troop, even though the former actions tend to reinforce solitude while the latter strengthen gregariousness. (Cf. Hinde, 1974.) The notion in question is elucidated by

DEFINITION 4.48 An animal engages in *social behavior* iff it acts on, or is acted upon by, other individuals of the same genus.

The condition of congenericity excludes preying and flight from predators from the social behavior repertoire but on the other hand includes reproductive behavior. Needless to say, social behavior may be automatic, in particular genetically programmed, as is the case with insects.

Although not all animals are gregarious or social (i.e. live in communities), all are social in the above sense, if not actually at least potentially. We assume therefore

POSTULATE 4.23 The behavioral repertoire of every animal includes types (patterns) of social behavior.

Like all behavior, social behavior is controlled both internally – in particular by the NES – and externally. For example, aggressiveness depends upon endogenous variables such as certain hormones (in particular testosterones) and the excitation of certain brain systems (e.g. the anterior hypothalamus), as well as upon exogenous variables, such as scarcity of food, cold, and crowding. Consequently biological determinism, which holds that all social behavior is determined exclusively by endogenous variables, holds only for animals whose social behavior repertoire is inherited, hence stereotyped. The social behavior of bees and ants, the ritualized courtship of birds, and the patterns of aggression and submission of cichlid fish and even of wild dogs, seem to be of that kind. On the other hand primates seem to inherit only those patterns of social behavior that concern reproduction and the care of the young: all others are learned. What primates do inherit is (*a*) the *need* for social intercourse and (*b*) the *capacity* to form and change rules of social behavior, and to form and undo social groups.

Animals get together to form families or communities, temporary or permanent, loosely or tightly knit, when such behavior improves their lot. The simplest such system is the family, and the simplest family is the mother–infant couple. The root of such a system is of course the care of the young. Note that families, though formed naturally – i.e. on the strength of natural bonds – are not biosystems. Only the components of a family are alive, i.e. metabolize, are capable of reproducing, etc. A family is a social system. We shall make a detailed study of social systems in the next chapter, but will offer now a definition that will be useful for our sociobiological considerations:

DEFINITION 4.49 A system σ is a *sociosystem* (or *social system*, or *social group*) iff

(i) the *composition* of σ is a set of animals of the same order (not necessarily the same species);

(ii) the *environment* of σ is the set of things, other than the components of σ, that act on or are acted on by the latter;

(iii) the *structure* of σ is the social behavior repertoire of the members of σ.

Every society is a sociosystem but the converse is not true: industrial plants, schools, and clubs are sociosystems but not societies. A society is a self-reliant sociosystem:

DEFINTION 4.50 A sociosystem is a *society* iff it is self-sufficient [i.e. does not depend entirely upon other sociosystems].

This definition will be refined in the next chapter.

5.4. *The Root of Social Cohesion*

What elicits the formation of a social system and keeps it together despite the somewhat divergent interests of its members? Our answer is this: rearing of the young in the case of the family, sharing (or participating) in the case of other sociosystems, and social pressure (peaceful or forceful) in all.

Cooperation need not be conscious: it can be automatic, as among social insects. Animals can cooperate either in their own individual interest (as is the case with sexual partners and with partners in a foraging expedition) or in the interest of their group – as in the case of an ant colony or a wolf pack. The general concept is elucidated by

DEFINITION 4.51 If a and b are animals, then a and b *cooperate* with one another iff the social behavior of each is valuable to the other or to a third animal.

When cooperation concerns things (e.g. goods) of some kind, it is called *sharing*, and *participation* when it concerns activities. More explicitly:

DEFINITION 4.52 Let σ be a social system with composition $\mathscr{C}(\sigma)$, environment $\mathscr{E}(\sigma)$, and structure $\mathscr{S}(\sigma)$. Further, let $T \subset \mathscr{C}(\sigma) \cup \mathscr{E}(\sigma)$ be a set of members of σ or of items in the environment of σ, and $A \subset \mathscr{S}(\sigma)$ a kind of activity in σ. Then for any component x of σ,

(i) x *shares* in T iff x cooperates with other member(s) of σ in acting on items in T;

(ii) x *participates* in A iff x cooperates with other member(s) of σ in doing A.

Now our hypothesis that association is based on cooperation:

POSTULATE 4.24 A set of conspecific animals forms a social system if and only if each of them cooperates with some other members of the same set [i.e. shares in the resources of the society or participates in some of its activities].

The early evolutionists stressed competition at the expense of cooperation, and both Freud and Lorenz regarded aggression as instinctive and pervasive. We have learned since that the struggle for life involves both. For example, flocking and the formation of colonies are effective defenses against predation (Hamilton, 1971). Group grazing, hunting in packs, and active defense involve cooperation – not to speak of higher forms of sociality such as division of labor and games. No cooperation, no sociosystems.

Only some forms of cooperation are purposeful, even fewer conscious. Purposeful cooperation deserves a name of its own:

DEFINITION 4.52 For any animals a and b,

(i) a behaves *altruistically* (or with *solidarity*) towards b iff a engages in purposeful social behavior that may be of value to b though not directly or immediately to a;

(ii) a and b are *reciprocally altruistic* (or *solidary*) iff a is altruistic towards b and conversely.

A rough measure of solidarity is this. Let a and b be animals, and $N(b)$ the set of needs of b (e.g. food, shelter, or care), and $G(a, b)$ the set of items a is prepared to give b. (Both $N(b)$ and $G(a, b)$ are taken during the same period of time.) Then the solidarity of a towards b is $G(a, b) \cap N(b)$. A quantitation of this concept is the following. *The degree of solidarity* of a towards b (over the same period) is

$$s(a, b) = |G(a, b) \cap N(b)| / |N(b)|.$$

(The more you give needed items the more solidary your behavior. Items that are not needed do not count.) The reciprocal solidarity between a and b is of course $s(a, b) + s(b, a)$. This measures the total number of items exchanged between a and b without regard to their value.

A more precise concept is easily constructed by assigning a value to each item needed and got. Such value need not be biological or psychological: in the case of higher animals it can be social, i.e. the items being traded off may benefit the social group as a whole. If only for this reason one can hardly subscribe to the cost-benefit analysis of altruism championed by Hamilton (1964) and Wilson (1975). This rests on the assumption that genes are so smart and species-conscious that they know it is good business for an individual to lay down its life for three or more sibilings, for in this case the chances are that its full genetic equipment (or rather a replica of it) will be saved for the benefit of posterity. Altruism requires a highly developed CNS capable of becoming aware that a conspecific is in need. Skipping levels, the way some sociobiologists do, won't do. (See Ruse (1979) for a balanced assessment of sociobiology.)

In addition to the inter individual actions that keep sociosystems going, there are the group influences on the individual that prevent it from rending the social fabric. Social pressure is not a mysterious action of the whole on the part but rather the action (direct or indirect) of the various components of a social group on those that deviate from the norm or mode. One way of elucidating the concept is by means of

DEFINITION 4.54 Let F be a function representing a property of members of a social system σ, and assume that the distribution of F in $\mathscr{C}(\sigma)$ is bell-shaped with average $\bar{F}$ and scatter d. Then, for any member x of σ,

(i) x *conforms* with respect to F iff $|F(x) - \bar{F}| < d$;

(ii) x *deviates* with respect to F iff x does not conform with respect to F.

We shall assume the obvious, namely that every social group contains some deviants, and that the group pressure on an individual increases with its deviance:

POSTULATE 4.25 In every social system (a) there are deviants in some respects, and (b) some of the fellow members of the deviant components subject them to a pressure that is a monotonically increasing function of the deviance.

In most vertebrate societies social structure has a rather weak reaction on individual behavior. Indeed flocks and troops of various kinds split and even dissolve altogether under altered environmental conditions. In hierarchical primate societies, on the other hand, the over-all social structure seems to be overpowering as long as the social system lasts. This makes for social stability – as well as for the occasional rebellion. The deviant, underprivileged, or just young, are often outcast. In a few cases they may find a more favorable habitat offering new opportunities and new challenges. In other cases the deviants may join forces to attempt a restructuring of their group. In any case, just as genic variety is a motor of biological evolution, so social deviance is a motor of social evolution.

5.5. *Communication*

All social animals, and many which are not, can exchange information of some kind, and some of them using a great variety of means. We shall adopt some of the standard concepts used in describing animal communication (Morris, 1938; Smith, 1965). They are all included in

DEFINITION 4.55 (i) An *animal signal* is a physical process executed or controlled by an animal, perceptible to other animals and capable of altering their behavior;

(ii) the *message* carried by an animal signal is an encoded representation of events in the CNS of the signalling individual;

(iii) the *significance* of a message to a recipient is the change in the latter's CNS caused by the signal carrying the message;

(iv) an animal *understands* a message if the events triggered in its CNS by the carrier signal are similar to those in the animal that emitted the signal;

(v) two animals *communicate* iff they understand the messages of the signals they exchange.

Remark 1 An animal signal may be blunt, as in pushing, or subtle, as in singing. *Remark 2* Communication is always among animals, though it is not always direct. For example, we do not communicate with computers but through them. *Remark 3* Although communication can be valuable (or disvaluable), it is not necessarily purposeful. For example, the signals exchanged by social insects, being genetically programmed, are not purposeful even though they are functional (biovaluable). *Remark 4* The psychoneural concept of significance (Definition 4.55 (iii)) differs from both the behavioral notion (Morris, 1938; Paivio, 1971) and the semantical one (Vol. 2, Ch. 7).

Not every set of signals is a language. For example, the set of human gestures is not a language. A language is a system of learned signals and moreover one that allows an animal to compose an unlimited number of distinct messages. A single signal suffices to build a language if allowed to concatenate with itself; on the other hand bees communicate without having a language: they do not learn their signals from scratch and they cannot generate an unlimited number of signals out of a few simple ones. The following convention summarizes all this:

DEFINITION 4.56 Let S be a finite nonempty set, and $\circ$ a binary operation in S. Further, call $\mathscr{L} = \langle S^*, \circ \rangle$ the free semigroup on S [i.e. the set of concatenates of members of S]. Then $\mathscr{L}$ is a *language* iff

(i) S^* is a set of learned animal signals;

(ii) there is at least one animal capable of understanding some of the messages carried by members of S^*.

Remark 1 This definition allows for private languages such as those deaf children can invent (Goldin-Meadow *et al.*, 1977). Incidentally, Wittgenstein and his followers have denied that private languages could exist. *Remark 2* Chimpanzees can learn certain sign languages (Gardner and Gardner, 1969). This finding has weakened the hypothesis that language is exclusively human – weakened rather than refuted, for the chimpanzees in question were taught a man-made language, namely the American Sign Language or ASL, whereas humans *create* languages. Moreover, although ASL has as much expressive power as any natural language, chimpanzees cannot make full use of it because they cannot think all human thoughts. *Remark 3* By using borrowed languages, chimpanzees can express thoughts of their own and thus refute the dualist's contention that it is "really senseless to conjecture that animals *may* have thoughts" (Malcolm, 1973). We are now sure, on the basis of experiments with teaching artificial languages to chimpanzees, that animals can think. *Remark 4* Because chimpanzees can be trained to communicate not only by means of ASL but also with the help of plastic chips (Premack, 1971) and via computers (Rumbaugh *et al.*, 1973), man can no longer be defined as *the talking animal*. On the other hand we can evidently postulate that humans (and chimpanzees and possibly other animals as well) have an innate linguistic ability or potentiality or disposition. This is not to say that man is born with a knowledge of certain basic structural features common to all natural languages. What we are born with is the vocal tract, the Wernicke and Broca areas, and subsidiary neural systems – as well as a social environment, carrier of a tradition, that stimulates the acquisition and development of verbal languages.

5.6. *Protoeconomy, Protoculture, Protopolity*

In the next chapter we shall argue that every human society, no matter how primitive or evolved, is composed of four main subsystems: kinship, economy, culture, and polity. We submit that all four are already found *in nuce* in some prehuman societies.

That prehuman societies have a kinship system, i.e. a social network based on reproduction, seems obvious. Nor is there any doubt that some have also an economic system, i.e. a social network based on work. As for politics in the large sense of management of social activities, it is well known than in certain animal societies law and order, as well as defense, are well organized. (Cf. Gray Eaton (1976) for the case of the Japanese macaque society.) Less well known, but true nonetheless is that some prehuman societies have a culture of sorts, i.e. a system whose members engage in social activities that are mental rather than biological, productive, or managerial. The patterns of such activities can moreover be transmitted throughout the community as well as handed over to the next generation – i.e. traditions are formed, spread, and are kept. Such a transmission is effected by imitation and occasionally also by teaching, hence with the help of signals. This has been found to be the case among Japanese macaques (Kawai, 1965) and chimpanzees (Menzel *et al.*, 1972). In both cases the initiation of new traditions called for some bold individuals – creative deviants – that sought newness rather than fearing it.

Because economic, cultural and political systems are already found in prehuman societies, although in a primitive form, we should characterize the corresponding concepts:

DEFINITION 4.57 If σ is an animal society, then

(i) the *economy* of σ is the subsystem of σ whose members engage in the active and organized transformation of the environment of σ;

(ii) the *culture* of σ is the subsystem of σ whose members engage in mental activities that control (or are controlled by) some of the activities of other members of σ;

(iii) the *polity* (or *political system*) of σ is the subsystem of σ whose members control (or are controlled by) the social behavior of other members of σ.

A particular case is that of tripartite societies composed of absolutely creative animals, i.e. animals capable of discovering and inventing certain things before any other animals (Definition 4.41). We submit that this is the case of all humans, from *Homo erectus* through *habilis* and *sapiens* to *sapiens sapiens*, i.e. us. That is, we assume

POSTULATE 4.26 An animal society is *human* iff

(i) some of its members are absolutely creative, and

(ii) it is composed of an economy, a culture, and a polity.

Remark 1 We are defining humanness in psychological and sociological rather than purely biological terms. In this we follow the tradition of anthropology and prehistory, and reject Neo-Darwinism. *Remark 2* Clause (i) suggests that the economy, the culture and the polity, far from being rigid, are plastic – as plastic as their members. Man is not only an economic, cultural and political animal, but also one that can alter very quickly almost any feature of this triad without having to wait for genic mutations or environmental cataclysms: man is the supreme creator and destroyer.

We now have all the preliminary material necessary to build an ontology of human society – a task for the next chapter.

6. CONCLUDING REMARKS

Wind is the motion of air: there is no wind apart from moving air. Likewise there is no metabolism over and above metabolizing systems, or social change over and above social systems. In every factual science, states are states of concrete entities, and events are changes of state of such objects. It is only in traditional psychology and philosophy of mind that talk of behavior independent of behaving animals, and of mentation alongside or above the brain, is tolerated, nay encouraged by the dualist myth. This separation between the mind and that which does the minding – between function and organ – has kept psychology and the philosophy of mind far apart from neurobiology and has prevented a full utilization of the scientific approach to the mind-body problem. The philosophy of psychophysical dualism is responsible for the backward state of research into that problem. Anyone interested in the progress of the science of mind must therefore reject the prehistoric myth of the autonomous mind.

The rejection of psychophysical dualism does not force one to adopt eliminative materialism in either of its versions – i.e. that mind and brain are identical, that there is no mind, or that all things have mental capacities. (This last version is of course indistinguishable from panpsychism or animism.) Psychobiology suggests not just psychoneural monism – the strict identity of mental events with brain events – but also emergentism, i.e. the thesis that mentality is an emergent property possessed only by animals endowed with a very complex plastic nervous system. This ability confers its owners such decisive advantages, and is related to so many other properties and laws (physiological, psychological, and social), that one is justified in asserting that the organisms endowed with it constitute a level of their own – that of *psychosystems*. However, this is not saying that *minds* constitute a level of their own, and this simply because there are no disembodied (or even embodied) minds, but only *minding bodies*. (The dual notions of embodiment and disembodiment are rather vague and typically dualistic, hence – *pace* Margolis (1978) – have no place in a science-oriented ontology.)

To repeat the same idea in different words: One can hold that the mental is emergent relative to the merely physical without reifying the former. That is, one can hold that the mind is not an entity composed of lower level things – let alone a thing composed of no entities whatever – but a collection of functions (activities, processes) of certain neural systems, that individual neurons do not possess. And so emergentist (or systemist) materialism, unlike eliminative materialism, is seen to be compatible with overall pluralism, or the ontology that proclaims the qualitative variety and mutability of reality. (Cf. Vol. 3, Ch. 5.)

Our espousing emergentist (or systemist) materialism does not entail claiming that it has already solved the mind-body problem. It has and it won't, for emergentist materialism is a philosophy providing only a scaffolding or general framework for the detailed scientific investigation of the many problems one lumps under the rubric 'the mind-body problem'. It behoves neuroscientists and psychologists to attack these problems – as scientists not as amateur philosophers let alone theologians. (Likewise emergentist materialism provides only a general framework for the detailed investigation of the problems of inanimate matter and of life, both of which are just as bottomless as the problem of mind.)

However, philosophy is far from behaving as a nonparticipant spectator of the scientific investigation of the mind-body problem. The dualist philosophy of mind has for centuries actively blocked the scientific approach to the problem. On the other hand emergentist (or systemist) materialism contributes to the scientific investigation of it by (*a*) dispelling confusions, (*b*) exposing myths, (*c*) suggesting that all problems about mental states or processes be construed as problems concerning brain functions, and therefore (*d*) encouraging the building of neural models of mental functions. As a matter of fact emergentist (systemist) materialism is the philosophy of mind behind physiological psychology, psychochemistry, and neurology. It is the only one that has been fruitful, that does not support quixotic reductionism, and that is compatible with a systemist and dynamicist ontology as well as with all the sciences.

CHAPTER 5

SOCIETY

This chapter is devoted to human society. Whereas to biologism human society is just one more animal society, according to spiritualism it is nothing of the sort because it is guided by ideas and values. We take the alternative view that human society is an animal society with many and remarkable novel properties, only a few of which it shares with animal protoeconomies, protocultures, and protopolities. We assume that man is neither an animal at the mercy of its genetic makeup and its environment, nor a free spiritual being akin to divinity: man is, instead, the primate that works and strives to know, that builds, maintains, and transforms social organizations far beyond the call of the gene or the environment, and that creates artistic, technological, and intellectual cultures, and also plays. Man is *faber* and *sapiens, oeconomicus* and *politicus, artifex* and *ludens*.

No aspect of mankind can be properly understood unless one keeps in mind the fivefold character of human nature – biological, psychological, economic, cultural, and political. The very existence of human society – of its economy, culture, and polity – has biological roots that cannot be cut, and psychological constraints that cannot easily be bent. There is no philosophizing on an empty stomach and, from a certain point on, no subsistence without new thinking, cooperating, and organizing. The biological, psychological, economic, cultural and political aspects hang together.

Human society is composed of animals but is not rigidly determined by biological factors, as shown by the variety of types of society. Moreover social behavior reacts on the animal basis by selecting socially or culturally valuable traits. This process may even oppose that of natural selection, as when the handicapped are protected. Thus social selection modifies genetic composition. For example whereas urbanization and crowding select resistance to stress and foul air, and perhaps also pushiness and resistance to bureaucacy, it abates other features. Such a social selection of biological and psychological traits modifies the genetic composition of the population, a change that has in turn social consequences. In sum there is a tight relationship between the biological and the social.

Consequently mankind does not evolve exactly as other species do: the theory of evolution does not apply to it nowadays the way it did a couple of million years ago. In fact we should distinguish three modes of evolution: "blind", animal, and social. "Blind" evolution is evolution by mutation, recombination and natural selection; it is the way molecules and plants evolve. Animal evolution combines genic change with behavior: the environment selects not only phenotypes but also behavior patterns, and animals adapt not only phylogenetically but also ontogenetically by learning adaptive behavior. Finally social evolution proceeds not only by genic change and behavioral change, but also by social creation and selection.

In all three cases the individual proposes and its environment decides – metaphorically speaking. But in the cases of animal and social evolution the individual learns and the environment, rather than being given to it, is partly chosen by and and partly even remade by it. And in the case of social evolution there are not only biological and psychological factors but also economic, cultural, and political ones. Moreover, thanks to language and social life, some of the learned skills and some of the products of human activity are transmitted to the new generations: here we can speak of the (social) inheritance of acquired characters. But even this inheritance is plastic, for we are free either to disregard or to enrich our heritage.

In the present chapter we shall examine the social aspect of man, centering our attention on social relations and the resulting social structure. Our aim will be to clarify and interrelate some of the key concepts of the social sciences – those which are so basic that no particular social science deigns to clarify them. We shall do it from a point of view that supersedes not only biologism and spiritualism, but also individualism and holism. This alternative is of course the systemist one. By regarding human society as a concrete system of a particular kind, and by analyzing the various subsystems and supersystems of society, we hope to attain some clarity about a traditionally obscure subject. Needless to say, our discourse will be extremely general: it will aim at some of the so-called cultural universals. Therefore we shall not be able to explain or predict any specific social events. We shall at the most supply a scaffolding for building specific theories in the social sciences.

1. HUMAN SOCIETY

1.1. *Definitions*

Our Postulate 4.26 characterizes a human society as an animal society containing some absolutely creative members, and composed of an economy, a culture, and a polity. That postulate served the purpose of defining (implicitly) humanness rather than human society. We shall now define the latter in some detail, assuming that biology, psychology and psysical anthropology can take care of the biological characteristics of *Homo*.

We submit that the chief characteristic of human society, in contradistinction to other animal societies, is that it possesses jointly all the following properties:

(i) some members of every human society do *labor*, i.e. engage habitually in the deliberate transformation of part of their environment (*homo faber*);

(ii) the workers use *tools* made in standardized patterns and work with them according to rules or *techniques* that they have invented, improved, or learned;

(iii) some members of every human society *manage* (direct or control or contribute to either directing or controlling) the activities of others: they organize labor or games, learning or fighting;

(iv) some members of every human community engage habitually (though perhaps not exclusively) in *cultural activities* – cave painting, story telling, tool designing, healing, reckoning, teaching skills, advancing knowledge, etc. (*homo culturifex*);

(v) all members of every human society devote some time to playing – sex, dancing, play-acting, racing, playing hide and seek, games of chance, etc. (*homo ludens*);

(vi) all members of every human society *communicate* with some other members of the same society or of other societies by means of standardized signs, in particular a language;

(vii) every member of every human community *shares* information, services or goods with some other members of the same community;

(viii) all the members of every human society *learn* attitudes, skills and information not only from their parents but also from other members of their community (by imitation or formal teaching);

(ix) every human society is divided into *social groups*, such as families and occupational groups;

(x) every human society lasts as long as every member of it *participates* to some extent in several social activities and benefits from such participation.

Surely other animals too work – e.g. bees and beavers. But they make no tools. Still others use natural tools, such as branches, stones or thorns, but only occasionally and in any case not fashioned by them or at least not to standardized patterns. Again, some birds and mammals play, and still others have evolved coded sign systems (though not languages proper). However, only humans organize (manage) some of their own activities, *and* teach behavior patterns and feeling and thinking patterns, *and* create novel objects (artifacts, works of art, organizations, words, etc.). In sum only humans possess, socially if not individually, *all ten* properties listed above.

The preceding characterization of human society, though sketchy, suffices for the moment. We proceed to express it in terms of concepts elucidated in the previous chapters:

DEFINITION 5.1 A concrete system σ is a *human society* (or *community*) iff σ is an animal society such that

(i) the *composition* $\mathscr{C}(\sigma)$ of σ is included in the set of human beings (i.e. the genus *Homo*, not necessarily our own species);

(ii) the *environment* $\mathscr{E}(\sigma)$ of σ contains some of the items necessary for the survival of some of the components of σ;

(iii) the *structure* $\mathscr{S}(\sigma)$ of σ equals the disjoint union of two sets of relations, S and T – called *social relations* and *transformation relations* respectively – such that

(a) S includes the relation of biological descent and all other (kinship) relations derived from it;

(b) every member of the set S of social relations is accompanied by communication (information flow);

(c) S includes the relations of sharing (goods and services) and participating (in social activities);

(d) S includes a nonvoid set $M \subset S$, called *management*, such that every member of M is a relation in some Cartesian power of the membership $\mathscr{C}(\sigma)$, and representing some habitual actions of members of σ upon fellow members, consisting in controlling certain relations in S (e.g. courtship and learning) or in T (e.g. hunting and ploughing);

(e) the set T of transformation relations includes a nonempty set $W \subset T$, called *work*, such that every element of W is a relation from a subset of $(\mathscr{C}(\sigma))^p \times (\mathscr{E}(\sigma))^q$, with $p, q \geq 1$, into a nonvoid proper subset A of $\mathscr{E}(\sigma)$, representing the habitual transformation, by some members of σ, of certain things in $\mathscr{E}(\sigma)$ (e.g. iron ore) into things in A (e.g. iron tools) called *artifacts*;

(iv) the social relations S and the transformation relations T generate equivalence relations inducing partitions of the membership of σ into social groups (social differentiation);

(v) σ is self-supporting as long as it lasts, i.e. is capable of satisfying its needs by work.

According to this definition not every bunch of humans, not even every sociosystem (Definition 4.49), constitutes a human society: to qualify as such a collection of humans must share an environment, transform the latter deliberately (by work, not just by wearing it out), hold social relations and communicate among themselves, be divided into social groups, and constitute a self-reliant unit. Thus a factory, a school and an army, though subsystems of a society and therefore sociosystems, are not societies. Further, the social relations among the members of the society must include man-transforming relations M such as educating, organizing and coercing, and they must be set up and maintained with the help of communication.

So much for a quick characterization of the concrete systems we call 'human societies', that make up the membership of the set of all human societies – past, present, and future. Let us emphasize that they are systems, not just either things or aggregates of such – let alone supraindividual entities of an immaterial nature. Note also that, although societies are composed of organisms, they have no organismic properties: they do not synthesize proteins, reproduce by cell division or conjugation, etc. In fact societies have bulk or systemic properties, some of which are peculiar to them:

DEFINITION 5.2 Let P be a property. Then

(i) P is a *societal property* iff there exists (now or at any other time) a human society possessing P;

(ii) if P is a societal property of some society σ, then P is a *resultant* property of σ iff P is also a property of some components of σ;

(iii) if P is a societal property of some society σ, then P is an *emergent* property of σ iff P is not a resultant property.

Example 1 The total food consumption per year of a given society is a resultant property of it for being the mere additive aggregation of the individual consumptions. Not so the total energy consumption, let alone the total production, for they are social. Likewise the family structure, the political organization, and the cohesiveness of a community are emergent societal properties. *Example 2* All members of a society possess properties that are neither individual nor societal, such as shyness and conformity (or its dual, deviance). Take the latter, analyzed by Definition 4.54. In it 'F' represents a property of an individual, the average $\bar{F}$ a resultant property of a sociosystem, and the standard deviation d of the distribution of F values in the society represents an emergent property of the latter. (So much so that if d is small we can say that the sociosystem is conformist.)

Our first hypothesis is that the concept of an emergent societal property is not idle:

POSTULATE 5.1 Every society has emergent properties.

We hasten to note, though, that 'emergent' does not mean "unexplainable" or "unpredictable". Firstly, because "emergence" is an ontological not an epistemological category. Secondly because it is the business of science not only to acknowledge emergence but also to incorporate it into theories and thus render it understandable and sometimes also predictable.

1.2. *Subsystem, Supersystem, Specific Function*

Every society, no matter how primitive, is composed of various subsystems. And almost every contemporary human society is a subsystem of a more inclusive social system, such as a province, a nation, a regional system, or the world system. So, we need

DEFINITION 5.3 Let σ be a human society. Then

(i) any social system σ' such that σ' is a subsystem of σ (i.e. $\sigma' \precsim \sigma$) is a *social subsystem* of σ;

(ii) any society σ' such that σ is a subsystem of σ' (i.e. $\sigma \precsim \sigma'$) is a *social supersystem* of σ.

Example The health system of a modern community is a subsystem of it characterized by the following coordinates:

Composition: physicians, nurses, administrators, clerks, janitors, inpatients, out-patients, etc.;

Immediate environment: hospital buildings, furniture, instruments, ambulances, families of the members of the system, etc.;

Structure: relations of diagnosing and prognosticating, medicating and operating, advising or otherwise acting on patients, keeping their families informed, organizing work, teaching, and supplies, keeping the wards clean, etc.

Not every component of a social system is a social subsystem of it. Thus the persons composing a social system are biosystems not sociosystems. Nor is every subset of the membership (composition) of a social system a social system itself. For example the employees of an industrial plant and the students of a school need not be social systems. It is only when organized, and acting as a whole in certain respects (e.g. going on strike), that a bunch of persons becomes a system. *A fortiori* the working class and the student population of a society may not constitute systems.

Consider now a homogeneous set of subsystems of a given society, such as the collection of all farms, all refuse collection and disposal systems, or all schools. Such sets are not systems but *kinds* of system. An arbitrary such set will be called an F-system, where F is short for the function(s) characterizing the member subsystems in contradistinction to others. More precisely, we put forth

DEFINITION 5.4 Let σ be a human society and call

$$S(\sigma) = \{\sigma' \in \Sigma \mid \sigma' \lesssim \sigma\}$$

the collection of all social subsystems of σ. Further, let F be a certain set of social relations or transformation relations, i.e. $F \subset S \cup T$, and let

$$F(\sigma) = \{\sigma' \in S(\sigma) \mid F \subset \mathscr{S}(\sigma')\} \subset S(\sigma)$$

be the collection of subsystems of σ where the F relations hold.
Then

(i) $F(\sigma)$ is called the *F-sector* of σ;

(ii) $G(\sigma) = \bigcap_{\sigma' \in S(\sigma)} \mathscr{S}(\sigma')$ is (are) called the *generic function*(s) of the subsystems of σ;

(iii) $F_s(\sigma) = \bigcup_{\sigma' \in F(\sigma)} \mathscr{S}(\sigma') - G(\sigma)$ is (are) called the *specific function(s)* of the F-sector of σ.

To put this definition to work think, e.g., of the retail, the school, the manufacturing, and the postal sectors of a community: see Figure 5.1.

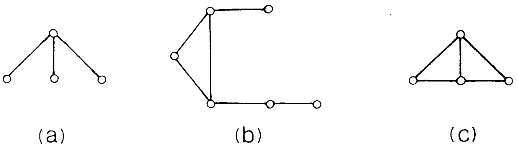


Fig. 5.1. Three sectors: (*a*) school system (connected indirectly via school board) or retail marketing system (connected indirectly via wholesale distributor); (*b*) manufacturing system (where the inputs to some subsystems are outputs to others); (*c*) postal network (controlled by central post office).

In all these cases the subsystems concerned perform certain peculiar functions that set them apart from the subsystems in other sectors; but they also share a number of functions proper to all social subsystems, such as consuming or transforming energy, producing waste products, and communicating with other subsystems of the community.

Primitive societies have relatively few sectors – e.g. the domestic work sector, the hunting sector, the manufacturing sector, the cultural sector (which is actually a system), and a few others. As society evolves some sectors become extinct while new ones emerge. Industrial societies are characterized by the large number and size of their sectors. They are also characterized by an increasing systematization of their sectors, i.e. an increasing coupling among the members of their sectors. Thus in a monopolistic economy every economic unit is a member of a tightly knit system. The general concept is elucidated by

DEFINITION 5.5 The F-sector $F(\sigma)$ of a society σ is the *F-system* of σ iff the membership of $F(\sigma)$ constitutes the composition of a subsystem of σ.

Finally we assume

POSTULATE 5.2 Every human society has a number of sectors. Moreover (*a*) every member of a human society belongs to at least two sectors of it and no individual belongs to all sectors at the same time, and (*b*) some sectors of every human society are systems.

Remark 1 In other words, there is some division of labor in every society, however primitive; yet the division of labor in human societies is never as thorough as in insect societies. *Remark 2* The above axiom, together with the epistemological postulate of knowability, entails the methodological theorem that every society can be ananalyzed into a number of sectors and, in particular, subsystems. What is nowadays called *large systems analysis* is mainly such an analysis, or recognition and characterization of the sectors and subsystems of a given society – such as, e.g., the transportation system of a city, and the marketing system of a region.

1.3. *Institution*

The word ‘institution’ is ambiguous: sometimes it denotes an individual sociosystem of a certain kind, such as a given school, and at other times it denotes a set of social actions and norms, such as marriage and religious cult. We are interested here in yet another acceptation of the word 'institution', namely the one used in speaking of the Family, the Mail, the Bureaucracy, or the Law.

Social philosophers are divided over the nature of institutions, hence over the proper way of studying them. The two classical doctrines in this field are holism and individualism. The holist regards an institution such as the Law as a disembodied corpus of rules or laws that individuals enforce, obey, or violate. According to this view the institution hovers above the persons and moreover is superior to or more valuable than them. No wonder holism, in some from or other, is inherent in totalitarian ideologies. The individualist rejects such an abstraction and regards the Law as the set of judges, lawyers and policemen – with total neglect of the subjects of the Law, namely the common citizens. When confronted with a sentence such as 'The Law compelled X to do Y', the individualist translates it into 'Judge (or policeman) U forced X to do Y'. But of course this is a poor substitute for the original sentence, as the person U in charge of enforcing the law behaves as he does because of what he is, namely a member of a certain legal system with a code of legal procedure. Different individuals might have behaved in roughly the same way. On the other hand the same individuals, in a different legal system, might behave differently. In conclusion, the legal system in question is neither the set including the individuals U and X nor a set of supraindividual rules. Hence although both holism and individualism have their points, neither is competent to account for the behavior of the individual in society. Holism fails because it postulates totalities transcending the components, and individualism fails because it ignores the totalities.

Our systemist view is as follows. A legal system is a subsystem of some society, i.e. it is a part of a human community set off by certain peculiar social relations of the M type, such as those of judging and passing sentence. And what is called 'the Law', with an awesome capital, is the collection of all legal systems. In general, we shall construe an institution as the set of all the F-sectors of all societies, regardless of their spatial and temporal propinquity. Thus the collection of all school sectors is called Education, the set of all trade unions Organized Labor, the set of all postal systems the Mail, and so on. (Recall Definition 5.4 of an F-sector and Definition 5.5 of an F-system.) We make then

DEFINITION 5.6 Let $\Sigma = \{\sigma_1, \sigma_2, \dots, \sigma_m\}$ be a set of human societies, and call F_{ik} the kth sector of society σ_i, i.e.

$$F_{ik} =_{df} = F_k(\sigma_i) = \{\sigma' \in S(\sigma_i) \mid F_k \subset \mathscr{S}(\sigma')\},$$

where F_k designates a certain set of social or of transformation relations. Then the family of F-sectors

$$\mathscr{F}_k = \{F_{ik} \mid F_{ik} \text{ is the } k\text{th sector of } \sigma_i \;\&\; \sigma_i \in \Sigma \;\&\; 1 \leq i \leq m\}$$

is called the *F-institution* of Σ.

In our construal, then, an institution is neither an individual sociosystem nor a Platonic set of rules, norms or goals governing people from above. But, since an institution is a *family of sets* of concrete things, it is tolerably abstract. (A sheer abstraction is one that, unlike an $\mathscr{F}_k$, does not boil down to concrete particulars. Sheer abstractions belong in logic and pure mathematics but are intolerable everywhere else.) To endow such a set of sets with properties of concrete individuals, such as life (or worse: an eternal life), a mind of its own (or worse: a deranged mind), aggressiveness, or the possibility of conflicting with similar types (e.g. other institutions), is to indulge in reification.

Note that not everything that passes for an institution in ordinary language qualifies as such according to our definition. For instance Work, War, Marriage, and Money do not because they are not families of sets of social subsystems.

It might be objected that our definition of an institution misses an essential ingredient, namely the goals and norms of the institution. Thus when applied to the case of the Law regarded as an institution, our convention seems to make no room for the codes of law and legal procedure that regulate the relationships at work in any legal system. The part such rules play is so important that some social philosophers and scientists have gone to the point of identifying an institution with the set of its norms or rules. This, though exaggerated and definitely Platonic, does contain a grain of truth: it is obvious that, unless certain rules are observed, the institution (or rather each of the subsystems forming it) decays or even becomes extinct. However, it does not follow that a set of institutional rules (e.g. the bylaws of an association) is an Idea dangling above the corresponding social subsystems, let alone identical with them.

The institutional rules are just the way the corresponding subsystems are thought (rightly or wrongly) to function optimally in the interest of society or of some social group. If preferred, those rules are prescriptions to operate the subsystems in an efficient manner – i.e. to attain the goals of its members, or rather of those whom the system is supposed to serve.

Undoubtedly such institutional rules can be rendered explicit for the most part, just as the rules for doing properly some job or playing some game can be formulated explicitly up to a point. But the writing down of a rule does not confer autonomous existence upon it: this may only suggest, wrongly, that it has one – perhaps because, when inscribed on a clay tablet, a stone slab, or a papyrus, it can outlast its author. The upright and merciless judge enforces rules just as the foreman has an operating manual observed.

The actions abiding by the rules of an institutional game may be called 'rule directed' as long as we do not suggest thereby the idealist thesis that rules exist independently of the rulers, the rule enforcers, and the ruled – anymore than the laws of nature are separate from the things satisfying them. The rules of social behavior are inherent in the very relationships holding among the members of the system in question. (Much the same holds for any deviations from that behavior: the delinquent may be an outlaw in the sense that he does not observe the code of law, but he is as much a member of the legal system as the judge – so much so that there is no need for a legal system where there is no occasion for delinquency.) In short the institutional rules, and also the patterns of breaking such rules and the rules for punishing these violations, are incorporated into the very network of social relationships that hold the system together. And it is not "society" that sanctions them or punishes any deviations from them, as the holist would have it, but rather some members of society, namely those in charge of regulating the functioning of the social system. By conceiving of institutional rules as patterns of the optimal or at least desired functioning of social systems, it becomes easier to understand how those rules originated, where they fail, and in what respects they could be improved upon if not repealed altogether.

So much for a preliminary conceptual framework. Let us now proceed to analyze more closely some of the typical concepts of the social sciences, starting with that of work.

2. SOCIAL SUBSYSTEMS AND SUPERSYSTEMS

2.1. *Work*

Work, one of the peculiarities of human behavior, is a kind of activity – not just an action but a regular one. An activity, whether performed by animal or by machine, may be defined as a patterned sequence of actions aiming at some preset goal. Thus, far from being erratic, an activity proceeds in such a way that, normally, the difference between attainment and target is gradually decreased (Wiener, 1948, Ch. IV). See Figure 5.2. We compress these remarks into

DEFINITION 5.7 A system performs an *activity* iff it executes a sequence of goal-directed actions.

For a system to perform an activity it is necessary and sufficient that it be a control system and that it, or whatever controls it, have a definite goal. In the case of work, the worker himself, together with his tools, constitutes a control system.

Unlike machines, humans are versatile or multipurpose systems: they can perform activities of very many kinds. This multiplicity can be represented as follows. Call ${}^{\sigma}A_{mn}$ the action performed by system σ on the *m*th aspect (property) of the *n*th thing in the environment of σ. Then the total activity of σ is displayed by its activity matrix ${}^{\sigma}A = \|{}^{\sigma}A_{mn}\|$. This is an $M \times N$ matrix, where M is the number of properties and N the number of things influenced by σ. An empty entry corresponds to a nonexisting feature of the corresponding thing. If an entire community, or a sector of it, with membership S, is engaged in a set of activities, we may define the aggregate activity of the community or sector as the matrix

$$A = \bigcup_{x \in S} {}^{x}A,$$

every entry of which is the set theoretic union of the partial entries with the same indices.

Work is a special kind of activity, namely one thought to be useful. We define it as follows:

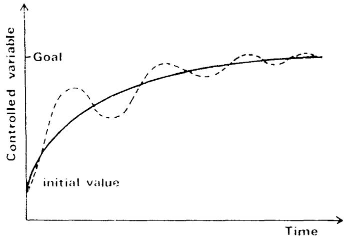


Fig. 5.2. A system attaining asymptotically a goal state in either of two typical ways: (*a*) monotonically (full line) or (*b*) by trial and error (dotted line). Note in the latter case the phenomenon of overshooting (momentarily overreaching the goal).

DEFINITION 5.8 If x is a system, then x *works* iff there exists a thing y and an animal z (equal to or different from x) such that

(i) x exerts an activity on y (in a way other than by eating y) and by so doing produces an appreciable change in the state of y, and

(ii) the thing y in the new state brought about by the activity of x is of value to z.

We distinguish three kinds of work, namely labor (L), cultural work (K), and managing (M). Labor aims at transforming nonhuman things into more valuable things; cultural work aims ultimately at modifying brain functions; and managing consists in controlling both labor and cultural work. See Figure 5.3. Let us first examine labor.

2.2. *Labor*

We start off by making

DEFINITION 5.9 Let σ be a society with composition $\mathscr{C}(\sigma)$, environment

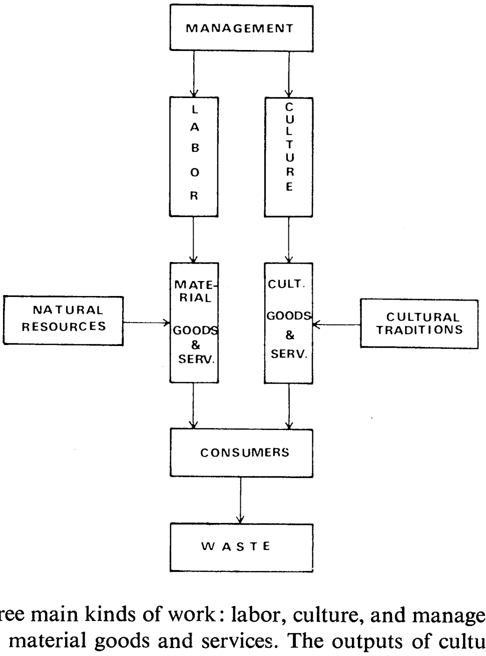

MANAGEMENT
LABOR
CULTURE
NATURAL RESOURCES
MATERIAL GOODS & SERV.
CULT. GOODS & SERV.
CULTURAL TRADITIONS
CONSUMERS
WASTE

Fig. 5.3. The three main kinds of work: labor, culture, and management of both. The outputs of labor: material goods and services. The outputs of cultural work: cultural goods (e.g. books) and services (e.g. education).

$\mathscr{E}(\sigma)$, and structure $\mathscr{S}(\sigma) = S \cup T$, and call L a subset of the transformation relations in σ. Then

(i) the element $L_i \in L$, with $1 \leq i \leq r$, is called the ith kind of *labor*, or *primary work*, done by members of σ, iff L_i is an activity performed on nonhuman things, the outcome of which is a set of things or a change of state in things, that is valuable to members of σ;

(ii) the domain of L_i is $(\mathscr{C}(\sigma))^p \times (\mathscr{E}(\sigma))^q$, where $p, q \geq 1$, and the range $A_i \subset \mathscr{E}(\sigma)$ of L_i is called the *product* of L_i, and each element of A_i an *artifact*;

(iii) the totality of products [goods and services] of the labor performed in σ, i.e. $\mathscr{A}(\sigma) = \bigcup_i A_i$, is called the *material production* of σ, whereas the complement $\mathscr{N}(\sigma) = \mathscr{E}(\sigma) - \mathscr{A}(\sigma)$ of $\mathscr{A}(\sigma)$ to the environment of σ is called the *natural environment* of σ;

(iv) the subset ${}^1P_i \subset \mathscr{C}(\sigma)$ occurring in the domain of the labor-relation L_i is called the set of *primary producers* of A_i, or the ith *labor force* of σ;

(v) the union ${}^1P = \bigcup_i {}^1P_i$ of primary producers of σ is the *labor force* or *primary manpower* of σ.

Remark 1 Our definition includes not only the production of material goods but also the performance of services. *Remark 2* It is not implied that labor be manual: it may well consist in controlling automated machinery. *Remark 3* Our concept of labor or primary work does not coincide with the standard concept of primary industry: the primary producers or laborers referred to above may be hunters or fishermen, farmers or industrial workers, miners or carpenters or repairmen. *Remark 4* Nor does our concept of an artifact coincide with the popular one: it is a technical one. Thus domestic bacteria, such as those used to make butter and cheese, as well as cultivated plants and domesticated animals, being products of artificial selection, are artifacts in our sense of the word. Machines are only a very special kind of artifact.

The set $L = \{L_i \mid 1 \leq i \leq r\}$ of kinds of labor cannot be empty for any society, under penalty of extinction. It may happen that a community subsists for a time on spoils of some kind, but even the most bountiful spoils of war become exhausted after a while. Moreover the set L must contain more than one member, for each kind of labor 'leads to' other kinds. Thus felling trees is done to employ timber for one or more purposes (activities), and it cannot be done efficiently, except by beavers, without axes or saws – which have to be manufactured. We compress these remarks into

POSTULATE 5.3 In any human society

(i) there is more than one kind of primary work, i.e. $| L | > 1$;

(ii) there are workers engaged (permanently or temporarily, exclusively or not) in every kind of work done in the society, i.e. $P_i \neq \varnothing$ for every $1 \leq i \leq r$.

Finally we make

DEFINITION 5.10 Let $\mathscr{E}(\sigma) = \mathscr{N}(\sigma) \cup \mathscr{A}(\sigma)$ be the immediate environment of a society with labor force ${}^1P$ and labor (primary work) relations in L. Then the *material force of production* of σ is the ordered triple $\langle {}^1P, \mathscr{N} \cup \mathscr{A}, L \rangle$.

Should any coordinate of this triple be empty, there would be no material force of production. In particular, an inactive (unemployed) labor force is as ineffectual as a nonexisting stock of natural resources. However, a scarcity in the latter can be compensated for by material production thanks to exchange. But someone, in some society or other, has got to produce food: there is no such thing as a postagricultural world society. Likewise there is no postindustrial society but only one in which most of the labor force is engaged in performing services rather than in agricultural or industrial work. No labor, no human society.

Let us move on to the other components of the manpower of any human society.

2.3. *Cultural and Managerial Work*

Let us now characterize the kind of work done by artists, bards, composers, performers, actors, story tellers, writers, teachers, scholars, scientists, engineers, and physicians of all descriptions, whether competent or not:

DEFINITION 5.11 Let σ be a society with composition $\mathscr{C}(\sigma)$ and social relations in S, and call $K \subset \mathscr{S}(\sigma)$ a proper subset of the social structure of σ. Then

(i) the element K_j of K, for $1 \leq j \leq s$, is called the *j*th kind of *cultural* (or *secondary*) work done by members of σ iff K_j is an activity capable of evoking feelings or thoughts, or supplying ideas intervening in the primary production of σ;

(ii) the subset ${}^2P_j \subset \mathscr{C}(\sigma)$ occurring in the domain of the cultural work relation K_j, is called the *j*th set of *cultural* (or *secondary*) producers of σ;

(iii) the union ${}^2P = \bigcup_j {}^2P_j$ is called the *cultural* (or *secondary*) *manpower* of σ.

Remark 1 Although cultural work aims at changing the way people feel or think, K is not disjoint from T. Indeed, all cultural work involves transforming some thing or other, if only the state of the surrounding air. Moreover there are useful works of art, such as artistically designed pieces of furniture, as well as useful works of science, such as scientifically controlled instruments. *Remark 2* According to clause (i), writing a book that may be published is doing a spot of cultural work; even writing or performing a piece of rock "music" is perpetrating some cultural work. Not so engaging in praying, dreaming, or transcendental meditation – which is sleeping anyhow. Likewise drawing the blueprint for a house or a machine or the pattern for a dress is engaging in cultural work, whereas drawing exclusively for one's pleasure, or humming a melody, are not. To be cultural an activity must be social or socializable as well as artistic or intellectual.

We shall return to the matter of cultural work in Sec.3.2. Let us now characterize the relations aimed at influencing social behavior in an immediate way, rather than through the subtle medium of culture:

DEFINITION 5.12 Let $M \subset S$ be a subset of the social relations in a human society σ. Then

(i) the element $M_k \in M$ is called the kth kind of *managerial* (or *organizing* or *tertiary*) *work* done by members of σ iff M_k is an activity contributing to controlling some primary or secondary work done in σ;

(ii) the subset ${}^3P_k \subset \mathscr{C}(\sigma)$, for $k = 1, 2, \ldots, t$, occurring in the domain of the managerial relation M_k, is called the kth set of *tertiary* (*managerial, organizing*) *workers* of σ;

(iii) the union ${}^3P = \bigcup_k {}^3P_k$ is called the *managerial force* (or *tertiary manpower*) of σ;

(iv) the subset ${}^3P_L$ of P^3 devoted to controlling the primary production of σ is called the *production management* of σ, and its complement ${}^3P_k = {}^3P - {}^3P_L$ the *culture management* of σ.

In other words, we call *managerial* all the kinds of work aimed at controlling production, whether agricultural or artistic, industrial or intellectual. Planners, plant managers, executives, members of socialist farm or factory councils, commune leaders, judges, bureaucrats, censors, and in general decision makers and transmitters of all kinds belong to management as construed above. Moreover one and the same individual may belong to two or more manpowers, as is often the case in primitive societies.

We assume

POSTULATE 5.4 Every human society has both a cultural manpower (i.e. ${}^2P \neq \varnothing$) and a managerial one (i.e. ${}^3P \neq \varnothing$).

From this platitude it follows that, under risk of destroying society, it is impossible to do entirely without either the cultural force or the managerial force. What one should aim for is maximizing the harmonious cooperation among all three work forces, i.e. easing the tensions and therefore eliminating the need for keeping repression forces. In particular, rather than try quixotically to eliminate management one should strive to make it work in the interest of material and cultural production rather than in its own interest. Such a blending of management with productive forces obtains in a well run modern industrial plant, marketing system, communications network, hospital, or school – in short wherever some people work according to the canons of advanced (scientific) technology. Such sociosystems deserve a name of their own:

DEFINITION 5.13 A system τ is a *technosystem* iff

(i) the composition of τ includes rational beings and artifacts;

(ii) the environment of τ includes components of a society;

(iii) the structure of τ includes production, maintenance or use of artifacts.

A modern American household full of gadgets does not qualify as a technosystem unless it is the seat of a family concern where intellectual work plays a significant role. On the other hand a modern defense system is a technosystem, for military experts are supposed to do some intellectual work as part of their duty.

The composition or population of any human society is made up of an active part – those who do work of some kind – and an inactive one. All those persons – housewives, shepherds, fishermen, hunters, craftsmen, industrial workers, white collar workers, physicians, engineers, poets, musicians, philosophers, etc. – who participate in the transformation of raw materials into artifacts, do artistic or intellectual work, or take part in organizing material or cultural production of goods or services, make up the active population or human resources. In short, we have

DEFINITION 5.14 Let σ be a human society. Then

(i) the *active population*, or *human resources*, of σ is the union of its manpowers: $A = {}^1P \cup {}^2P \cup {}^3P$;

(ii) the *inactive population* of σ is the complement of A to the membership of σ, i.e. $I = \mathscr{C}(\sigma) - A$.

2.4. *The Main Subsystems of a Human Society*

Neither the active population nor the inactive one constitute systems; nor are the subsets ${}^1P$, ${}^2P$, and ${}^3P$ systems. Thus the managerial force of a society is hardly a system: managers are components of either the economic, the cultural, or the political systems of a society. Each of these systems draws its components from all three manpowers: see Figure 5.4. Thus industry, trade and government employ artists, intellectuals and professionals of various descriptions; and both manual and clerical workers take part as auxiliaries in cultural production as well as in political management. We have then the division

$$ {}^iP = {}^iP_L \cup {}^iP_K \cup {}^iP_G \quad \text{for} \quad i = 1, 2, 3 $$

of each manpower into one part engaged in primary production, another engaged in cultural production, and another engaged in political control. Whereas the upper index i identifies the specific output of the sector, the lower indices name its various inputs.

Correspondingly we distinguish certain subsets of the collection S of social relations: we call S_L, S_K, and S_G respectively the social relations associated with primary production, cultural production, and political administration. Finally, we group similarly the work relations: we call

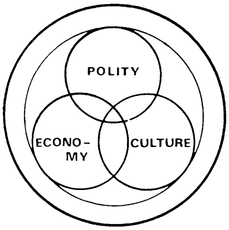


Fig. 5.4. The main subsystems of any society. Outer circle: whole society. Inner concentric circle: kinship system. Eccentric circles: *economy* (gathering, hunting, fishing, manufacturing, trading, financing); *culture* (artistic, technological, scientific, humanistic creation and their diffusion); and *polity* (government of all kinds, nongovernmental political activity, and the military). The overlaps suggest that some persons participate in two or even three systems. E.g. a scientist belongs to the cultural system by virtue of his research and teaching; to the economic system if he writes textbooks; and to the political system if he acts as consultant to the government.

L_K and L_G repectively the kinds of primary work done in cultural production and in government by members of the labor force; K_L and K_G the kinds of cultural work involved in primary production and in politics respectively; and M_L, M_K and M_G the kinds of management involved in primary production, cultural production, and politics respectively. We are now ready for

DEFINITION 5.15 Let σ be a society with labor force ${}^1P$, cultural force ${}^2P$, and managerial force ${}^3P$. Further, let the subscripts L, K and G identify whatever is associated with primary production, cultural production, and political administration respectively. Then

(i) the subsystem of σ represented by

$$\varepsilon_\sigma = \langle {}^1P_L \cup {}^2P_L \cup {}^3P_L, \mathscr{E}(\sigma), S_L \cup (L \cup K_L \cup M_L)\rangle$$

is the *economy* (or *economic system*) of σ;

(ii) the subsystem of σ represented by

$$\kappa_\sigma = \langle {}^1P_K \cup {}^2P_K \cup {}^3P_K, \quad \mathscr{E}(\sigma), S_K \cup (L_K \cup K \cup M_K)\rangle$$

is the *culture* (or *cultural system*) of σ;

(iii) the subsystem of σ represented by

$$\pi_\sigma = \langle {}^1P_G \cup {}^2P_G \cup {}^3P_G, \mathscr{E}(\sigma), S_G \cup (L_G \cup K_G \cup M_G)\rangle$$

is the *polity* (or *political system*) of σ;

(iv) ε_σ, κ_σ, and π_σ are the *artificial* main subsystems of σ.

The previous definition allows us to restate Postulate 4.26 in a more complete way:

POSTULATE 5.5 (i) Every human society is composed of four subsystems: the kinship, the economic, the cultural, and the political systems; (ii) every sociosystem in a human society is part of at least one of its four main subsystems; (iii) every individual in a society belongs to the composition of at least one of the four main subsystems of the society.

Remark 1 By conceiving of the economy, the culture and the polity as systems, one avoids the sterile philosophies of holism and of individualism. For the same reason one obtains the possibiity of modeling the entire economy (or culture or polity) of a society as a system with definite composition and structure – provided one includes in the latter the links of the economy (or the culture or the polity) with the other subsystems. *Remark 2* All three artificial subsystems share the same natural and artificial environment; in particular no culture and no polity can function in a physical vacuum. Hence extreme internalism (e.g. cultural idealism) is just as inadequate as extreme externalism (e.g. ecological or geographic determinism). *Remark 3* Each manpower is distributed among the three artificial subsystems. In particular, not even the most primitive economic system fails to employ some cultural workers and some organizers – even if they are at the same time primary producers – and not even the most anti-intellectual of political systems dispenses altogether with intellectuals, e.g. as planners and censors. *Remark 4* It is possible for one and the same individual to belong to more than one of the main subsystems. It is impossible for an individual to fail to belong to at least one of the main subsystems – i.e. complete marginality is impossible within a society. *Remark 5* A significant change in any of the three coordinates of each subsystem (membership, environment, structure) will affect the system as a whole. Examples: an increase, or decrease, in the percentage of skilled workers or of intellectuals; an increase, or decrease, in the consumption of energy *per capita*. When the change is both drastic and fast, it is called a *crisis* – of growth or of decline. *Remark 6* Because each of the three artificial subsystems is a subsystem of the same society, it is linked to the other two. Such couplings would be impossible if the systems in question (the economy, the culture, and the polity) were sets of individuals (individualism) or Platonic ideas (idealistic holism). *Remark 7* Disregarding or even minimizing the importance of any of the four main subsystems of a society leads to lopsided views, hence inadequate doctrines of political action. The same holds for neglect of the interactions among the subsystems.

2.5. *The Supersystems: Nation and World*

We regard a *nation* as a social system, or rather supersystem, composed of all the kinship, economic, cultural, and political systems coexisting and interacting within a territory. Unlike the city-states of Antiquity, modern nations are composed of several societies, each of which occupies a given geographical region within the national territory (which need not be in one piece). Each component society contributes its own kinship, economic, cultural, and political subsystem to the whole. The artificial subsystems are coupled to one another forming a national economy, a national cultural system, and a political system. None of these national systems need be homogeneous: on the contrary, variety makes for viability. (The ideology according to which homogeneity is a supreme desideratum is called *nationalism*, and is particularly vociferous in demanding racial purity, cultural homogeneity – i.e. poverty – and political uniformity – i.e. dictatorship.) If the integration among the various societies composing a nation is weak, then the nation is so only in name, as was the case with the Holy Roman Empire most of the time. If the integration is too strong, then the nation is cut off from other nations. The best coupling strength is middling.

The relations among or within the three main artificial subsystems of a nation are of the following kinds: (*a*) economic – e.g. society *X*'s supplying energy to society *Y*; (*b*) economicocultural – e.g. research scientists's developing a new variety of wheat; (*c*) economicopolitical (e.g. an economic group's lobbying for a bill favoring its interests); (*d*) cultural (e.g. intercity library loans); (*e*) politicocultural (e.g. a governmental agency's encouraging or obstructing technological innovation); (*f*) political (e.g. the armed forces seizing control of certain governmental agencies).

All of these relations are intersystemic. Yet in the final analysis they rest on person to person relationships, where the persons concerned act out their social roles. Thus when analyzing a social relation *S* between two subsystems of a nation we may ultimately explain *S* in a way such as this: System *b* bears relation *S* to system *c* if and only if there is an individual compoment *x* of *b* and an individual component *y* of *c*, such that *x* bears *S* to *y*, where *S* is an inter personal bond such as, e.g. writing a note asking for some item of property on loan. (Clearly the individuals concerned may be represented by robots, but the latter are built to act not in their own interest but in that of their master.) In other words, the relations in a supersystem such as a nation can be analyzed, at least in principle, in terms of interpersonal relations, as long as the roles of the persons are identified – roles which are not further reducible to properties of isolated persons. If such roles and the associated responsibilities are forgotten, i.e. if societal relations are mistaken for private relations, not only a faulty analysis results but a law may be broken. In short, although every social system is ultimately composed of individuals, these happen to engage in relations that are incomprehensible without the hypothesis that the system has supraindividual properties. (Recall Definition 5.2.)

Economists, sociologists, politologists, and social historians are becoming increasingly interested in whole regional blocs and international systems, thus recognizing that the nation is often either a subsystem or a supersystem, and in either case an inadequate unit of study. (The model for social history is Braudel's *La Méditerranée* (1949). A recent addition to the systemist approach is Wallerstein's *The Modern World-System* (1974).) Indeed there are no isolated nations left it there ever were. Every modern nation bears economic, cultural or political relations to other nations, and this on four levels: person to person, subsystem to subsystem, nation to nation, and international organization to international organization. (It goes without saying that all system to system relations are designed and implemented by individuals. However, these individuals can often be replaced by others enacting the same roles or functions. Not that their personalities are unimportat; only, these serve or disserve their roles or functions.) Let us take a quick look at international relations. (Cf. Deutsch, 1968; Fritsch, 1974; Bunge *et al.*, Eds., 1977.)

In addition to some 150 nations there are today about 10,000 international organizations. The latter include (*a*) international governmental organizations such as the World Health Organization, (*b*) international nongovernmental organizations such as the International Union for the History and Philosophy of Science, and (*c*) multinational corporations such as the General Motors Co. Since every other social system is either a subsystem of some international organizations or closely related to some of them, it would be utopian to plan anything on a national scale without taking the international context into account. This context is the *international system*, composed by all nations and all international organizations; its environment is no less than the biosphere; and its structure the set of relations (economic, cultural, political, economicopolitical, politicocultural, etc.) among its various components. These relations are rather entangled. For example, nation A trades with nation B and has close political ties with nation C, a rival of B; consequently the trade relations are affected by the political ones and reciprocally. Politologists tend to say that this is a case of relations among relations or cross relations. Since this notion is fuzzy, we prefer to analyze the same facts using our systemic framework, namely as follows.

Every international actor (nation, intergovernmental organization, multinational corporation, etc.) has three chief subsystems, every one of which can interact with any of the subsystems of the other two. See Figure 5.5. By completing this diagram we realize that two members of the international system can engage in 9 *sets* of binary relations, to wit: $E_1 - E_2$, $E_1 - C_2$, $E_1 - P_2$, $C_1 - E_2$, $C_1 - C_2$, $C_1 - P_2$, $P_1 - E_2$, $P_1 - C_2$, $P_1 - P_2$. For n members of the supersystem, the total number of sets of binary international relations is $N = (9/2)n \cdot (n - 1)$. For the entire interna-

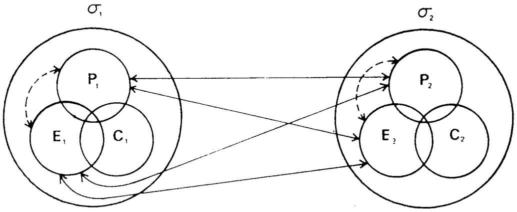


Fig. 5.5. Example of international economic and political relations between two nations σ_1 and σ_2. Note the two intranational relations besides the four international ones: the former are indicated by dotted lines.

tional system today, $n \cong 10{,}000$, whence $N \cong 450$ million sets of international relations. (The number computed on the assumption of the ghostly relations among relations is far more staggering, namely of the order of 10^{15}, or 7 orders of magnitude greater than ours.) This is a forbidding figure yet still modest by comparison with the number of pairwise relations among the atoms in a wedding ring, namely 10^{38}. The politologist can get around this hurdle just the way the physicist does, namely by taking into account only the relations among nearest neighbors. In this way whole blocs emerge, and then the relations among blocs can be studied – just as with any other supersystem analyzed into lower level subsystems. Such a systems analysis paves the way to the construction of theoretical (mathematical) models including precise functional relations among the variables. (Cf. Bunge, 1977h.)

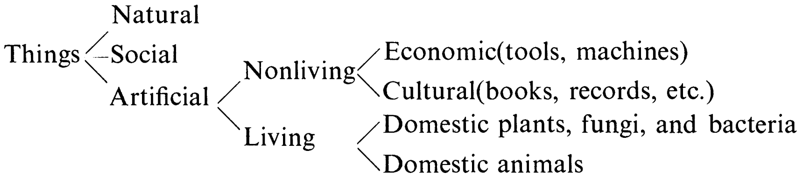
So much for a preliminary study of the main subsystems and supersystems of human society. Let us now take a closer look at each of the three artificial subsystems of every society.

3. ECONOMY, CULTURE, AND POLITY

3.1. *The Economic System*

According to Definition 5.15 the economic system of a society is composed of its primary workers (${}^1P_L$), cultural workers (${}^2P_L$), and managerial workers (${}^3P_L$). The primary workers are engaged in primary production (L), the cultural workers are in charge of the cultural input to primary production (K_L), and the managerial workers organize both (M_L). Thus the machinists, electricians, truck drivers and maintenance workers of an industrial plant belong to ${}^1P_L$; the engineers, quality controllers and draughtsmen, to ^2P_L; the executives, accountants and typists, to ^3P_L; the foremen often belong to all three sets but their specific function is as members of ^3P_L. All these workers share the same environment and bear (production) relations S_L among themselves in addition to whatever other relations they may hold. The input–output schema of an economic system is shown in Figure 5.6.

Of all the aspects of the economic system, those of greatest philosophical interest are perhaps technology and the artifacts it produces and employs. Both raise numerous philosophical problems of all kinds, hardly investigated (cf. Bunge, 1977g). Let us say a word or two about artifacts, i.e. things fashioned by rational beings with the help of some technique, whether primitive or advanced. An artifact is not just one more thing but a thing belonging to a kind that did not exist before the emergence of man or some other rational being. Artifacts may be regarded as constituting a whole new level of reality, namely the *artiphysis*. The set of all concrete things may now be classed as follows:


Things
Natural
Social
Artificial
Nonliving
Economic(tools, machines)
Cultural(books, records, etc.)
Living
Domestic plants, fungi, and bacteria
Domestic animals

Far from being contemptible entities whereof a philosopher does not deign to speak, artifacts are at the heart of human society and have properties absent from natural things. To understand this point it may help to

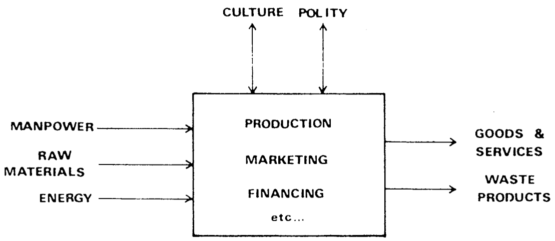

CULTURE
POLITY
MANPOWER
RAW MATERIALS
ENERGY
PRODUCTION
MARKETING
FINANCING
etc...
GOODS & SERVICES
WASTE PRODUCTS

Fig. 5.6. Schema of the economic system. Inputs: natural resources (raw materials and energy) and human resources. Outputs: goods, services, and waste products. Some of the outputs are fed back into the system (e.g. as tools). Controls: culture (mainly techniques and advanced technology), and politics.

recall the various items entering the production, servicing, use, and circulation of artifacts. They are chiefly the following: (*a*) the idea or design of the artifacts; (*b*) the idea of the possible use(s) of the artifact; (*c*) a plan or design for the manufacture of the artifact; (*d*) the raw material and the energy required to produce the artifact, (*e*) the actual production process; (*f*) the marketing of the finished product; (*g*) the control of the entire process. Of all these seven ingredients only one, the fourth, is supplied by nature. Because all the others are ultimately in the hands or brains of people, the artiphysis is totally at their mercy. Men, or similar social animals, create it and can destroy it. Artifacts have no will of their own and, indeed, no existence of their own: even the most sophisticated among them are but tools. And technology has no autonomy either: it is cultivated only in societies that care for it. A band of technoclasts could kill every technologist and destroy every artifact on Earth in a few days. So, to claim that technology and the set of artifacts have a life and a momentum of their own, that they cannot be stopped, is bizarre to say the least.

We summarize the foregoing in

POSTULATE 5.6 (i) Every human society has members able to design, manufacture, repair, or use artifacts; (ii) every nonliving artifact [tool or machine] serves only the purpose(s) of some animal or other.

It follows that no machine, not even a highly evolved robot, has a purpose of its own: whatever the goal seeking mechanisms built into it, they have been designed and ultimately manufactured by rational beings in their own interests. Hence machines stand no chance of taking over unless of course programmed to do so. It follows also that, when a nonliving artifact imitates rational animals in certain regards, it does so by proxy not spontaneously. A further consequence is that artifacts can neither evaluate nor think on their own. That they do not evaluate is obvious since they have no goals other than those built into them by their designers. (This vicarious evaluation capacity may reside in the hardware, as when a machine is designed to seek or to avoid certain stimuli or else to rank them. Or in the software, as when a machine is programmed to weigh apples or alternative routes for mail delivery. In either case delegated evaluation may trigger a vicarious decision – e.g. that of discarding an underweight apple or an inconvenient route. When speaking of 'the vicarious abilities of machines' the technologist underlines 'abilities' whereas the philosopher emphasizes 'vicarious' – unless of course he takes the former's word.)

As for our claim that machines cannot think, it is quite obvious in the light of the psychobiological view on thought as an activity of certain neural systems (Definition 4.35). Here again machines, unlike rational beings, have no purposes – in particular cognitive goals – of their own, so they can properly be said to act by delegation. Surely machines can help us solve certain problems, but they cannot discover any because formulating a problem is not a rule directed activity and, as long as we do not master a rule, we cannot design or program a computer to follow such a rule. (This is also why there can be no self-programming machines: if a machine were to have problems of its own we would quickly scrap it.) Also, machines can be used to process the assumptions of certain theories, but they do not help create them – again for the same reason, namely because theory construction is not a rule directed activity but a thoroughly creative one. In short tools, however marvelous, are nothing but tools.

3.2. *The Cultural System*

In Definition 5.15 we characterized the cultural system of a society as being composed not only of cultural workers (${}^2P_K$) but also of primary workers (${}^1P_K$) and managerial workers (${}^3P_K$). The former are of course in charge of cultural production (K), whereas the other two groups are attached to them either in an ancillary capacity, as in the case of the maintenance workers at a university, or as organizers, as in the case of university administrators. Another example: the technicians and the typists working for an experimental biologist belong in ${}^1P_K$, while the director of the scientist's laboratory is in ${}^3P_K$, and occasionally also in ${}^2P_K$.

All the members of a cultural system share the same environment and are linked by certain relations S_K. But not all of the ingredients of a cultural system are on the same footing. Its pivots are: (*a*) the workers engaged in direct cultural pursuits, i.e. ${}^2P_K$; (*b*) the part of the natural or social environment that is the object of contemplation, study or rational action; (*c*) the part of the stock of artifacts composed of cultural artifacts such as pictures, books, microscopes, and the like, and (*d*) cultural work proper – e.g. research, writing, and instructing.

Ours is a restricted concept of a cultural system. A wider concept includes the entire cultural force of the society, i.e. not only the persons working for the sake of culture but also those engaged in an auxiliary capacity by the economic and political systems, such as the technologists in a quality control unit or the sociologists in a census bureau. However, this cultural supersystem is much less tightly knit or integrated. Thus the chemical engineer working in an oil refinery is more strongly attached to the economic system than to the cultural system *stricto sensu* – unless of course he is a research chemist, in which case he will belong to both the economic and the cultural systems, and may accordingly be torn by a conflict of loyalties. In either case the input–output schema of the cultural system is sketched in Figure 5.7.

This schema serves as a reminder that, whether in the narrow or in the comprehensive construal, a cultural system is a *concrete* system – not just "a body of meanings, values and rules". More precisely, a culture is a concrete system composed of living rational beings engaged in activities of various kinds, some of which go beyond the call of biological need, and all of which are social because they draw from tradition and influence social behavior. I.e. the creation and the diffusion of culture are social processes, shaped by society and in turn having an influence on the economy and the polity.

Our concept of a culture as a concrete system is at variance with the idealistic notion of a disembodied system of ideas, values, and rules, which is the concept favored by most anthropologists. Actually this other notion is that of a *type* of culture rather than of a particular culture. Thus where a cultural system (in our sense) includes certain actions, such as writing and inventing, a culture in the anthropological sense includes only the patterns of such actions – e.g. the rules that are supposed to regulate those actions. Nor does our concept of culture coincide with that of an ideal superstructure – made up of ideas, values, and norms – mounted on the material structure (economy and polity). Unlike these dualistic concepts of culture, ours is monistic.

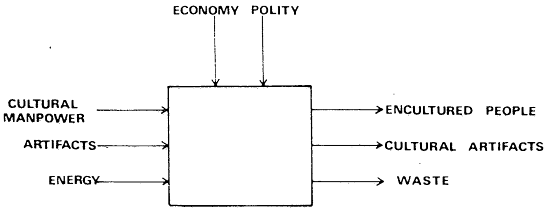


Fig. 5.7. Input-output model of the cultural system. Inputs: manpower (pivotal and auxiliary), artifacts (paint brushes, paper, pencils, books, archives, etc.), and energy (light, heat, etc.). Outputs: cultural artifacts (musical scores, sculptures, journals, etc.), people transformed by the latter, and waste products. Some of the outputs are fed back into the system. Controls: economy and polity.

The cultural systems of primitive societies are at least as monolithic as their economic and political systems, in the sense that they are not composed of subsystems. The emergence of civilization was accompanied by an explosion in the division of work, in particular cultural work. The single shaman was succeeded by a cohort of healers, priests, bards, teachers, painters, and later on skilled craftsmen and even scribes. Accordingly the cultural system split into a number of subsystems or subcultures – the religious system (centered around temples), the educational system (centered around schools), and so on. These various subcultures complement one another in certain respects but are at odds in others, if only because they all compete for a finite number of people as well as for a finite amount of material resources. In any case they interact. The mutual actions among the main subsystem of contemporary culture are shown in Figure 5.8.

Both the internal mechanisms of each cultural subsystem and its interactions with other members of the cultural system drive its changes, however slight and slow. The successive states of a cultural system constitute the *cultural evolution* (or *history*) of the society concerned. Cultural evolution modulates biological evolution and both proceed by selection. But, unlike biological evolution, cultural evolution is not purposeless. Cultural innovation and diffusion, be it technical or artistic, scientific or ideological, is partly random, partly deliberate.

A flourishing culture is one that teems with newness – valuable newness

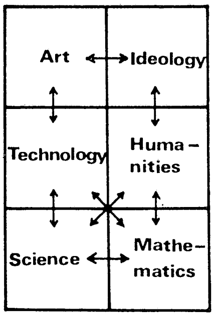


Fig. 5.8. The contemporary cultural system and its subsystems. Each box represents both creation and diffusion. The arrows represent both information flows and controls. Art and ideology, which have the most numerous memberships in traditional societies, have the fewest information links with the remaining subcultures. (From Bunge, 1977g.)

not just oddity – and is free to develop it without too much interference from the economic and political systems. A declining culture is one that ceases to value discovery and invention, and takes refuge in lamaistic repetition or in withdrawal from reality. The management of a cultural system – the politics of culture – can encourage creativity, channel it, or nip every creative bud. A dead culture is still a culture and one that can remain in that state for a long time. But because of the functional relationships among the three main artificial subsystems of every society, any major change in the economy or in the polity is bound to have cultural repercussions.

The cultural system, then, coevolves with the economic and political systems. However, the modes of evolution of these subsystems may be quite different, and neither is merely an image of the other. Thus cultural growth is compatible with zero economic growth and political stagnation, provided the political system does not interfere destructively with cultural evolution. Moreover, while economic development is limited by natural resources, there are no such limits to cultural growth: the more we know the more fresh problems we can pose and solve.

Yet there are of course economic and political constraints on cultural evolution. For one thing any given society can support only so many professional composers, mathematicians, philosophers, and so on. But at least here, unlike the case of the economy, a solution is in sight, namely automation and the amateur pursuit of cultural interests. In fact the former is freeing many hands and brains from routine work, and the good society should turn that added freedom into leisure and creation instead of unemployment. A revaluation of culture should encourage people to engage in cultural work for the fun of it rather than primarily as a source of income. Everyone should be able to master whatever cultural skill he or she likes, and to practice it during the increasingly long leisure hours rendered possible by automation. This need not encourage substandard cultural production. On the contrary, it is the increasing substitution of extrinsic motivation (pay, promotion, fame) for intrinsic motivation that is lowering the quality of cultural work. (For the negative effects of extrinsic rewards on children see Lepper *et al.*, 1973.)

To conclude. Our view of culture is at variance with idealism ("The spirit is the driving force") as well as with economism ("Culture is but a byproduct of the economy"). If every culture is a subsystem of a society then it has its own dynamics, hence some degree of autonomy, and it also interacts with the other subsystems, namely the economy and the polity.

Therefore it is neither totally free and omnipotent nor totally enslaved and powerless. Just as some members of the economic system exercise economic power, and some members of the political system wield political power, so some members of the cultural system exercise cultural power, particularly if they are entrenched in certain cultural subsystems, whether private or governmental. For example, the school system of a community exerts some cultural influence on all the inhabitants of the community – often as strong an influence as that exerted in the past by organized religion. This influence is not restricted to purely cultural matters. A creative cultural organization may study, discuss, and even propose and circulate blueprints for economic and political action. Such proposals do not amount to action but may arouse and guide it. After all people are being moved by thinking individuals all the time. Why let obsolete ideologies do all the pushing when it is possible to draw rational and feasible blueprints – i.e. to ground social action on the social sciences?

3.3. *The Political System*

The political system of a society is the subsystem of the latter that controls (to some extent) economic and cultural work as well as other types of social behavior. This control is ultimately exercised by individuals (political leaders and bureaucrats) in the interests of a few or of the majority, arbitrarily or according to rule. There is no such thing as "a government of laws, not of men" – the modern substitute for the fiction of government in the name of a deity.

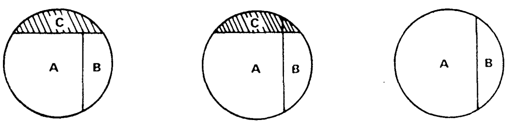
Since even infants are affected by certain political actions, one might regard the entire population of a society as composing its political body. However, for practical reasons it is convenient to restrict the composition of the political system to its actors or active members, namely the citizens with political rights or responsibilities: all others are politically marginal or passive. The structure of a political system is of course the set of social relations among the political actors, such that effect a political control. Every political *regime* (e.g. parliamentary democracy) is characterized by certain types of political relations.

The political system of a modern society, unlike that of a primitive society, has a distinguished subsystem, namely the *state* or *government*. In turn, the state of an advanced society may be analyzed into a number of subsystems, such as the defense system, the health system, the educational system, and so on. Some such subsystems constitute a relatively hard core ("technical", "nonpolitical", nonideological) that remains nearly invariant under changes in political leadership. For better or worse many such a governmental subsystem grows and strengthens to the point where it escapes almost all control. If it possesses physical force, as is the case with the army and the police, it may end up by controlling the entire political system and, through it, the whole society.

The differences among political regimes are differences in the distribution of political power, i.e. in the participation in political action. Consider a community formed by three social groups to be called A, B, and C. Assume further that A and B are mutually disjoint and that C is the political control group. Moreover all the C's have effective control: they cannot be outvoted on every important issue. Under these conditions there are exactly three possibilities: see Figure 5.9. The left circle illustrates the case of aristocratic (or theocratic) government, where the control class does not mesh with the other groups. The middle circle illustrates the various kinds of democracy (representative, participatory, etc.). The right circle illustrates anarchy: no control proper. The degrees or measures of the participation of A's and B's in C are respectively

$$\alpha = |A \cap C| / |C|, \qquad \beta = |B \cap C| / |C|.$$

In addition to such objective measures of political control, the politologist is interested in the manner in which the political system is "perceived" by people. That is, politics is not only a matter of objective political structure and power but also one of social psychology. This was seen by the founder of political science, in particular when he remarked that rebellions are caused either by the desire for equality or the desire for superiority (Aristotle, *Politica*, Bk. V, Ch. 2). Likewise Marx emphasized the



Nil participation:
No A's or B's participate in C:
$A \cap C = B \cap C = \emptyset$
$\alpha = \beta = 0$.

Partial participation
Some A's and B's participate in C:
$A \cap C \subset A, B \cap C \subset B$
$0 < \alpha, \beta < 1$.

Full participation
All A's and B's participate in C:
$C = A \cup B$
$\alpha + \beta = 1$.

Fig. 5.9. Three modes of political participation (or regime).

difference between the objective and the subjective conditions for a revolution. Let us take a quick look at two key concepts in political psychology: those of adhesiveness, and feeling of unfairness.

Whether or not a society is highly cohesive is of less consequence to its "survival" than the way that cohesiveness is "perceived". A social fabric with a low actual cohesiveness may be "seen" as highly cohesive (because e.g. of the repression of dissent); and conversely a highly cohesive society may be "seen" not to be so (because e.g. of a vocal nonconformist minority). The key concept here is that of attachment of an individual to a group or system. We may say that an individual feels attached to a social group or a social system if he believes that the benefits (material, cultural, emotional, etc.) he derives from sticking to it are greater than the price he has to pay for doing so. In other words, the *adherence* or *attachment* of individual x to a given group or system G_i is $A_i(x) = B_i(x) - C_i(x)$, where B and C are the benefit and cost functions respectively. There is positive attachment to G_i if the benefits outweigh the costs (i.e. $A_i(x) > 0$). Otherwise there is either rejection ($A_i(x) < 0$) by x of G_i or indifference ($A_i(x) = 0$) of x to G_i. The *total adherence* or *attachment* of individual x to a society σ formed by n social groups G_i is

$$A(x) = \sum_{i=1}^{n} A_i(x),$$

and the total *adherence* or *attachment of the membership* of σ to σ is

$$A(\sigma) = \sum_{x \in \mathscr{C}(\sigma)} A(x).$$

By choosing a suitable normalization factor, $A(\sigma)$ may be made to range, say, between -1 and 1. For example, the 1978 elections showed that the total adherence of the French people to their society is 0 (or $\frac{1}{2}$ on a [0, 1] scale).

Because adherence is psychosocial it can change swiftly in the light of information, misinformation, and criticism. When such means fail to enhance adherence, some structural social change is impending. Those in power can meet a decrease in popular attachment in either of two ways: with concessions or with coercion. The extent of either depends upon the imbalance between the actual adhesion value and the minimum value required for the social structure to remain unaltered.

Our second concept of interest is that of equity "perception". The political behavior of any political agent is partly determined by his "perception" of the equity or inequity of his station in society and his evaluation of the equity (or inequity) ideal. The imbalance between the two will contribute to his attachment or disaffection to the prevailing social structure. Therefore the social scientist is interested not only in actual equity but also in perceived equity, desired equity, and desirable or feasible equity. Let us take a quick look at these concepts.

Anything that can be distributed or allocated, from goods and services to leisure hours to political participation, can be assigned with various degrees of fairness or equity. Now, what is fair in one society may be unfair in another. Thus whereas in certain societies an average of 10 schooling years would be fair for all, in others this would constitute too heavy a burden, hence be unfair to the majority, and in still others it would be insufficient, hence just as unfair to the majority. This notion of an objectively desirable value of a property (such as the number of schooling years) must be distinguished from that of a subjectively desired value (desideratum). Whereas the latter may be unrealistic and incompatible with fairness, the former should in principle be determinable by studying the state of development and resources of the society. Such an objectively determinable value (or distribution, as the case may be) we shall call an *equity goal*, in contrast with the multiple *equity desiderata*.

We may assume that for each feature or variable there is one measure of equity or, dually, inequity. Whatever the feature may be, the objective inequity is some function f of the difference between the actual value X of the variable and its goal value G, i.e. $i(X) = f(X - G)$. *Example 1* The variable has a single value at a given time, as is the case with the mean number of schooling years. In this case a suitable inequity indicator is just the difference between G and X divided by the greater of the two. *Example 2* The feature in question is a distribution, such as that of arable land among the families of a society. If the goal is a straight line (e.g. the area of land is proportional to the number of families), then the Gini index is a suitable measure of inequity: see Figure 5.10. *Example 3* The goal distribution is symmetric – as is the case of the distribution of working hours *vs.* age. In this case inequity equals skewness, of which there are various measures.

The acuteness of the perception of inequity, as well as the desire for equity, depend on a number of factors in addition to actual inequity. One of them is social mobility. In a primitive society equity is taken for granted, whereas in a highly stratified society inequity may go just as un-

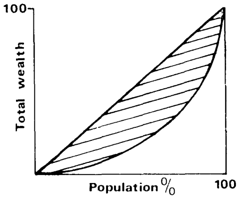


Fig. 5.10. Inequality in the distribution of wealth. The Gini index is the difference between the area under the straight line (equality) and the area under the actual curve.

questioned for being sanctified by ideology. On the other hand where social mobility is appreciable many persons feel status differences as iniquitous and also avoidable either by individual or group action. If the former proves ineffectual, the time is ripe for dissent and even sedition. We may then adopt the following version of the *Aristotle-de Tocqueville-Marx law*: Extreme inequity ends up by being perceived and negatively evaluated, that diminishes the attachment to the social structure, and this in turn leads to either indifference or active rebellion – in either case to a deterioration of the political system.

Needless to say, inequity may be economic, cultural, or political. Inequity felt in just one respect is neither necessary nor sufficient for a group to rebel against the social fabric. Any group that fits fairly well in the prevailing social (economic, cultural, and political) order – i.e. that is not marginal – is likely to be conservative even though that group may not be the principal beneficiary of that order. What do seem to be necessary and sufficient for members of a social group to rebel against the prevailing social order are (*a*) to feel being discriminated against in all three respects (economically, culturally, and politically), (*b*) to feel free from the traditions attached to the given social order (i.e. to be cultural not just economic or political outcasts), and (*c*) to be able to organize themselves into a system.

Finally we come to the notion of freedom. There are as many kinds of freedom as there are kinds of social behavior. Since human social activities occur in either of the three main artificial subsystems of society, we may group the species of freedom into three genera: economic, cultural, and political. Freedom to work (or to loaf) is economic freedom; freedom to learn is cultural freedom – and so is the freedom to remain ignorant; and freedom to participate in making decisions affecting the entire community is political freedom – and so is the freedom to remain politically inactive. Given the variety of freedoms and the fact that each has its opposite number, we should be wary of characterizing a society as free without further ado. We should specify in what respects and to what extent the members of a society are free or unfree. But such a specification presupposes an exact concept of freedom. Let us try and supply one, albeit one that is of interest to politology, namely that of *potestas agendi*, or freedom *to*, rather than that of freedom *from* (hunger, cold, fear, etc.). We may characterize the power to act by

DEFINITION 5.16 Let A be a set of alternatives, each consisting in engaging in an activity or in abstaining from it. Further, let x be an arbitrary member of a society σ. Then

(i) person x is *free* to do $y \in A$ in σ iff y is a possible choice for x & x can effectively do y & x values y & the benefit x expects to derive from doing y is greater than the cost to x of doing y;

(ii) society σ is *free in respect A* iff every member of σ is free to do any $y \in A$;

(iii) society σ is *free* iff σ is free in every respect, i.e. for any set A of alternatives (or choices).

Our definition of power or freedom to act rests on the notions of alternative or choice, of capacity or competence to perform an act, of value, and of cost and benefit – all of which are presumably elucidated in some discipline or other. It exhibits the dependence of freedom of action upon laws and circumstances: it construes freedom as depending on determinacy (whether causal, stochastic, or other) rather than on indeterminacy (hence lawlessness). In a lawless universe, in fact, there would be no freedom of any kind, since in order to be free do to something an agent must be able to count on certain regularities, at least those involving his possibilities, valuations, and actions.

We shall assume the rather platitudinous

POSTULATE 5.7 Every person and every society are free in some respects, none in all.

A completely free society – i.e. one in which every person were totally free – would be impossible, because the exercise of one person's freedoms is limited by that of his neighbors'. And a thoroughly enslaved society would be just as impossible, because every person needs to exercise some freedoms if only to survive and to contribute his or her share to society. Consequently neither extreme libertarianism nor extreme authoritarianism is a feasible blueprint for an enduring society – let alone an endearing one.

The matter of personal freedom constitutes the center of ideological debate in our time, yet it cannot be argued with profit on the ideological plane. On the other hand the systemic view of the political system can shed some light on that debate, namely as follows. A normal person is a component of some society or other. To survive, he or she must keep some measure of individuality or autonomy as well as engage in some synergic (harmonious) interactions with other components of the system, for otherwise the latter would break down. Indeed, exaggerating the autonomy of the individual would lead either to his or her breaking away from the system, or to subjecting the latter to the tyrannical control of the individual at the expense of the freedom of the other components. And overstressing the synergy of the whole may lead to depriving their components of creativity and thus to impoverishing society. So, if we care for both the system (not necessarily the "establishment") and the individual, we had better devise a political regime combining the good points of individualism with those of collectivism. But the matter of political regimes will have to be deferred to Vol. 7 of this *Treatise*.

4. SOCIAL STRUCTURE

4.1. *Root Social Relations*

We have drawn a distinction between a relation in general and a connection (or link, coupling or bond) in particular (Ch. 1, Sec. 1.2). In the case of social relations we distinguish social bonds, such as exchange relations, from nonbonding social relations such as that of belonging to the same occupational group. If a bond holds between two individuals, then at least one of them will behave differently from the way he or she would behave if not so bonded. On the other hand a nonbonding relation makes no difference to the relata. (Only the realization of belonging to the same social group may induce a feeling of either solidarity or hostility and thus influence social behavior. In the jargon of sociologists, the "perception" of a relation is bound to affect behavior.)

Among all the social bonds we now single out the root or basic ones, i.e.

those on which all the others depend – such as the kinship relations and the relations of economic, cultural, and political power. We may characterize them as follows:

DEFINITION 5.17 A *root* (or *basic*) social relation is a social relation that is
(i) bonding,
(ii) person to person (rather than system to system), and
(iii) an interaction (rather than a one-sided action).

We assume that all social relations are either basic or generated by relations of this kind:

POSTULATE 5.8 If σ is a society, then

(i) for any nonbonding person-to-person relation X in the structure of σ, there is a bonding person-to-person relation Y, in the same structure, such that, for any components a and b of σ, if Xab then there are components c and d of σ such that Ycd, where c and d are possibly equal to a and b respectively;

(ii) for any (bonding or nonbonding) system-to-system relation U in the family of subsystems of σ, there is a bonding person-to-person relation V in the structure of σ such that, for any subsystems x and y of σ, if Uxy, then there is at least one component a of x, and another b of y, such that Vab.

Example 1 Slave owners a and b, living in different places and times (hence not bonded), are socially equivalent ($\sim$) by virtue of being masters (M) of slaves (S): $a \sim b =_{df} (\exists x)(\exists y)(Sx \,\&\, Sy \,\&\, Max \,\&\, Mby)$. *Example 2* Factory a supplies concern b with merchandise (or goods) c, or $Sabc$ for short. This boils down to the following: members of a produce c, and members of a (the same or different) deliver c to members of b. In obvious symbols: $Sabc =_{df} (\exists x)(\exists y)(\exists z)(x \in \mathscr{C}(a) \,\&\, y \in \mathscr{C}(a) \,\&\, z \in \mathscr{C}(b) \,\&\, Pxc \,\&\, Dycz)$.

Social bonds may in turn be classed into voluntary and involuntary. Voluntary relations are those the individual can enter by choice – e.g. friendship, participation in community activities, and even – in industrial societies in times of full employment – job relations. Examples of involuntary social ties are the filial relation, the relation of being trained at a nursery school, the prisoner–captor relation, and the bondages of slavery and serfdom. We assume that, just as bonds determine nonbonding relations, so involuntary ties determine some of the voluntary ones. For example, being a prisoner, a serf, or a slave, determines most of the other relations a person can bear to other individuals. The mere existence and the basic character of involuntary relations suffices to brush aside the philosophy of voluntarism, according to which the individual will is the source of all behavior. The truth lies between the extremes of voluntarism and total determinism: man can act of his own free will (Theorem 4.3) but not always (Postulate 5.7).

Of all the involuntary social bonds family relations are the most basic because they are determined by reproduction, which is a biofunction. The family is thus situated at the biosocial interface. The root of the family is then the division of the sexes. In other words, the same-sex relation $\sim_s$ splits the total membership $\mathscr{C}(\sigma)$ of a society σ into two disjoint sets: males (M) and females (F). In short, $\mathscr{C}(\sigma)/\sim_s = \{M, F\}$. A human family of the nuclear or restricted kind may be defined as an aggregate of human coresidents among whom the relations of mate (or spouse) or of parenthood obtain. The relation μ of mate (or rather its graph) is included in the cartesian product $M \times F$. The relation π of being parent of somebody is a binary relation in $\mathscr{C}(\sigma)$. And the relation ρ of coresidence too may be construed as a binary relation in the same set. This is all we need to form

DEFINITION 5.18 Let σ be a society with sexual division $\mathscr{C}(\sigma)/\sim_s$, mate relationship μ, parenthood bond π, and coresidence tie ρ. The *nuclear family* $F(b)$ of the component b of σ is the sociosystem constituted by the mate(s) and the children of b who live with b:

$$\mathscr{C}(F(b)) = \{x \in \mathscr{C}(\sigma) \mid (\mu xb \vee \pi bx) \,\&\, \rho xb\},$$
$$\mathscr{E}(F(b)) \supset \mathscr{E}(\sigma) \,\&\, \mathscr{E}(F(b)) \cap \mathscr{C}(\sigma) \neq \varnothing,$$
$$\mathscr{S}(F(b)) = S \cup T, \quad \text{with } S = \{\mu, \pi, \rho\}.$$

A member of a nuclear family may have more than one spouse (as in the cases of bands, polygamy, and polyandry) provided all his or her mates are coresident. But each person belongs to exactly one nuclear family. (A childless widow constitutes a one person nuclear family but may belong to an extended family.) The set of nuclear families of a society is the composition of the nuclear kinship system of the society. That set is split into disjoint sets by the relation of belonging to the same family. The importance of this relation and of the partition it induces justifies making

DEFINITION 5.19 Let σ be a society. Then

(i) two components of σ are *family-wise equivalent* ($\sim_f$) iff they are components of the same family:

For any $x, y \in \mathscr{C}(\sigma)$, $x \sim_f y =_{df} (\exists z)\ (z \in \mathscr{C}(\sigma)\ \&\ x \in \mathscr{C}(F(z))\ \&\ y \in \mathscr{C}(F(z))\}$;

(ii) the *nuclear kinship system* of σ is the subsystem whose members are in some family or other, so that its membership is

$$\mathscr{C}_n(\sigma) = \{x \in \mathscr{C}(\sigma) \mid (\exists y)\ (y \in \mathscr{C}(\sigma)\ \&\ x \in \mathscr{C}(F(y))\};$$

(iii) the *family structure* of σ is the set of all the nuclear families of σ, i.e. the partition of $\mathscr{C}_n(\sigma)$ by the relation $\sim_f$ of family equivalence:

$$\mathscr{F}(\sigma) = \mathscr{C}_n(\sigma)/\sim_f.$$

Because of the biological root of the family, the family structure of a society constitutes its fundamental structure. So it is the basis of other social partitions. For example, a social class is a certain set of families, hence a subset of the family structure rather than a collection of individuals. So much for the so-called blood relations.

Other important social relations are those said to be "based on" work, i.e. those people enter by virtue of their work. Such are the "vertical" relations between a blue collar worker and a member of the management of his enterprise, and the "horizontal" relation between two workers on the same echelon in an organization. See Figure 5.11.

DEFINITION 5.20 Two persons hold a *bonding relation of production* iff either

(i) both individuals take part in the production of a set of products (goods or services) either as primary (L) or as secondary (K) workers; or

(ii) one of the individuals produces material or cultural goods while the other manages (controls, oversees) him or her; or

(iii) the two individuals comanage at least one producer.

In other words, the relations of production can be of coproduction, management, or comanagement. They are all, whether directly or indirectly, "based" on primary work insofar as, without the latter, there would be no

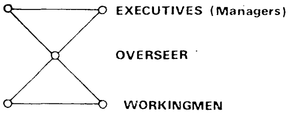


Fig. 5.11. Fragment of a particular web of social relations of production in an economic system.

production at all. And without the relations of production the social structure would be far poorer. Note, however, that not every relation of production is bonding. For example, the relation of belonging to the same occupational group, i.e. of doing work of the same type, is not bonding. It is only when individuals with the same occupation gather to form a guild, trade union, or some other organization "based on" work, that they enter into bonding relations. But these relations are no longer economic: they are political in our sense of the word, for their main impact is on social behavior rather than or production.

Whether economic, cultural, or political, social relations influence behavior. If such influence is decisive, we call it 'power' or 'control'. The social influences of all kinds fall under

DEFINITION 5.21 Let σ be a society where an activity of type $A \subset \mathscr{S}(\sigma)$ is performed. Then for any components x and y of σ, *x influences y in respect A* iff the activity of kind A performed by y while it communicates with x differs from that performed by y when it does not.

This definition makes room for indirect influences, i.e. those mediated by third persons. Thus even if a worker never sees a stockholder of his company, the two influence each other via the company management. (However, the influence relation we have just defined is not transitive – nor is it intransitive.)

Concerted influence is group pressure: this is not a mysterious influence exerted by a group as a totality, but the simultaneous and convergent influence of several individuals of group on a single individual. More precisely, we make

DEFINITION 5.22 Let $s \subset \mathscr{C}(\sigma)$ be a subset of the membership of a society σ. Then s exerts (group) *pressure* on a component x of σ in respect A iff each member of s influences x in respect A.

Power or *control* may be said to exist when influence is far stronger in one direction than in the opposite direction. It is possible to quantitate influence and power, hence to define a relation of power equivalence inducing the partition of the membership of a society into groups wielding the same power in some respect (Bunge, 1974d). However, we must rush on to other matters. The point of bringing in the above considerations is that, in order to disclose the social groups in a society, we must spot bonding social relations with the help of which (nonbonding) social equivalence relations can be defined, that induce the splitting of the membership of a society into homogeneous social groups. This is the method that we shall use to display the structure of a society.

4.2. *Social Structure*

Each social bond $B_k \in \mathbb{B} \subset S$ in a society generates a social equivalence relation $\sim_k$, which in turn induces a partition $\mathscr{P}_k = \mathscr{C}(\sigma)/\sim_k$ of the membership of the society σ into a family of homogeneous sets (social groups). There are as many such partitions as social equivalence relations. The collection of all such partitions we call the social structure of the society. More precisely, we propose

DEFINTION 5.23 Let σ be a society with composition $\mathscr{C}(\sigma)$ and call

$$\sim\; = \{\sim_k \mid \sim_k \subset \mathscr{C}(\sigma) \times \mathscr{C}(\sigma) \;\&\; \sim_k \text{ is an equivalence relation } \&\; 1 \leq k \leq n\}$$

the set of social equivalence relations definable on $\mathscr{C}(\sigma)$ on the basis of social bonds. Then

(i) the kth *primary social structure* of σ is the kth partition of $\mathscr{C}(\sigma)$, i.e. $\mathscr{P}_k(\sigma) = \mathscr{C}(\sigma)/\sim_k$, with $\sim_k \in \sim$;

(ii) the ith *social* group of the kth primary structure of (or *ik cell* of σ for short) is the ith member S_{ik} of

$$\mathscr{P}_k(\sigma) = \{S_{1k}, S_{2k}, \ldots, S_{ik}, \ldots, S_{mk}\}, \quad \text{with} \quad \bigcup_i S_{ik} = \mathscr{C}(\sigma) \text{ and } \bigcap_i S_{ik} = \varnothing;$$

(iii) the *overall primary social structure of* σ is the family of all the partial social structures of σ, i.e.

$$\mathscr{P}(\sigma) = \{\mathscr{P}_k \mid \mathscr{P}_k = \mathscr{C}(\sigma)/\sim_k \;\&\; \sim_k \in \sim \;\&\; 1 \leq k \leq n\}.$$

The overall social structure of a society is displayed by the matrix $\|S_{ik}\|$, every entry of which represents a social group or cell. (For the mathematics of set matrices see Bunge, (1974d).) The corresponding quantitative concept is got by just counting the population of each cell at a given time, i.e. by finding out how the total population $|\mathscr{C}(\sigma)| = N$ of the society is distributed among its various groups. Calling $N_{ik} = |S_{ik}|$ the population of cell S_{ik}, we can form the *population density matrix of* σ:

$$\|D_{ik}\| = (1/N)\,\|N_{ik}\|, \quad \text{with} \quad \Sigma_i N_{ik} = N.$$

A primary social structure is a partition of the membership of a society into groups or cells of individuals. But in some cases the groups themselves may band together to form larger groups or even systems. For example, the collection of families may be split into social classes, educational groups, income groups, political factions, and so on. Again, the collection of enterprises may be split into agricultural, industrial, and service enterprises – or into state owned, cooperative, and privately owned, and so on. Any such further partition of a primary partition will be called a secondary structure:

DEFINITION 5.24 Let $\mathscr{P}_k(\sigma) = \mathscr{C}(\sigma) / \sim_k$, with $\sim_k \in \sim$, be a primary social structure of a society σ. Then

(i) the *kmth secondary structure* of σ is

$$\mathscr{P}_{km}(\sigma) = (\mathscr{C}(\sigma) / \sim_k) / \sim_m, \quad \text{with} \quad \sim_m \in \sim \quad \text{and} \quad m \neq k;$$

(ii) the *overall secondary social structure* of σ is the totality of partial secondary structures of σ:

$$\mathscr{P}\mathscr{P}(\sigma) = \{\mathscr{P}_{km}(\sigma) \mid \mathscr{P}_{km}(\sigma) = (\mathscr{C}(\sigma) / \sim_k) / \sim_m \,\&\, \sim_m \in \sim \,\&\, m \neq k \,\&\, 1 \leq k, m \leq n\}.$$

Obviously not all such partitions are sociologically significant. For example, it makes no sense to group enterprises according to cultural background, or to partition the set of families into privately managed, cooperatively managed, and state managed. The sets resulting from such unnatural or insignificant partitions are either disregarded or set equal to the empty set. In particular, $\mathscr{P}_{kk}(\sigma) = \varnothing$. And, as with the case of the primary social structure, we can count items and build the density matrix of the overall secondary structure of a society.

Because both the primary and the secondary structures of a society are sets of sets, it might be wondered whether they stand for properties of the society. They do, for they may be regarded as values of property-representing functions, namely

$$\mathscr{P}_k \colon \Sigma \longrightarrow \{\mathscr{C}(\sigma) / \sim_k \mid \sigma \in \Sigma\} \quad \text{and}$$
$$\mathscr{P}_{km} \colon \Sigma \longrightarrow \{(\mathscr{C}(\sigma) / \sim_k) / \sim_m \mid \sigma \in \Sigma\},$$

where $\mathscr{C}(\sigma)$ is the membership of σ and Σ the totality of social systems. Furthermore, far from being arbitrary these functions occur in societal laws.

The concepts of overall primary and secondary social structures allow us to exactify the vague notion of social fabric: The *social fabric* of a society σ is the set of all primary and secondary social structures of σ, i.e.

$$\mathscr{P}(\sigma) \cup \mathscr{P}\mathscr{P}(\sigma).$$

Finally, one word about social class and class structure. Not all social groups are social classes, and therefore not all societies partitioned into groups need be stratified. A social class or caste is in a control relation with some other groups in the society. The control can be economic, political, military, administrative, or cultural. Such dominant groups originate in a variety of manners, from military conquest to gradual concentration of power. A typical case is that of medieval aristocracies. Small bands of soldiers, initially hired to defend a territory from marauders or from war lords, ended up by using military force to gain full economic and political control. What had been a horizontal division of society became a vertical one, i.e. a stratification or hierarchical order of social classes. This mechanism is quite general: the concentration of power in a social group turns it into a class. And, because power can be economic, cultural, or political, we need to distinguish three kinds of social class. That is, one and the same society may be divided into economic, cultural, and political classes. Occasionally these three partitions coincide, but in general they do not. For example, in a theocracy there are essentially two classes: the priestly caste, which wields total economic, political and economic power, and the laity. On the other hand in a contemporary Western democracy there are essentially two economic classes (the owners and the non-owners of the means of production), two cultural classes (the producers and the consumers of cultural goods and services), and – at least in theory – a single political class (the citizenry). We compress all this into

DEFINITION 5.25 Let σ be a society, and s and t two groups of σ. Further, let $\sim_A$ be a relation of social equivalence in respect A, where A stands for 'economy', 'culture', or 'polity'. Then

(i) group s *controls* group t in respect A iff every member of t is controlled in the respect A by some member(s) of s;

(ii) an *A-social class* is a social group that A-controls, or is A-controlled by, other social group(s) in the same society;

(iii) the *A-class structure* of σ is the partition of the composition of σ by $\sim_A$, i.e.

$$C_A(\sigma) = \mathscr{P}_A(\sigma) = \mathscr{C}(\sigma)/\sim_A;$$

(iv) the *class structure* of σ is the family of all partial class structures of σ, i.e.

$$C(\sigma) = \{\mathscr{P}_A(\sigma) \mid A = \text{economy, culture, or polity}\}.$$

If the quotient of $\mathscr{C}(\sigma)$ by $\sim_A$ fails to partition the composition of σ into disjoint subsets, then there is no A-control in the society, and thus the society is A-classless. In other words, σ is *classless* in the respect A iff $\mathscr{C}(\sigma)/\sim_A = \{\mathscr{C}(\sigma)\}$. It follows that an A-classless society is possible only if each of its components participates extensively in activities of kind A. The problem of participation will be addressed in the next subsection.

4.3. *Differentiation and Cohesion*

No society is homogeneous in all respects: every society is differentiated, i.e. split into social groups – if only sex and age groups. A natural though admittedly coarse measure of the difference between two social groups A and B is the numerosity of their symmetric difference $A \triangle B = (A \cap \bar{B}) \cup (\bar{A} \cap B)$. Hence we obtain a measure of the social differentiation (or variety) of a society by adding up the cardinalities of the pairwise differences among all its social cells, i.e. the $|S_{pq} \triangle S_{rs}|$. And, since it is convenient to have the resulting number comprised between zero and one, we divide the total by a suitable normalization factor. The end result (Bunge and García-Sucre, 1976) is

DEFINITION 5.26 Let σ be a society with social structure matrix $\|S_{ik}\|$, where $S_{ik} \in \mathscr{C}(\sigma)/\sim_k$ is the ith cell ensuing from the partition of $\mathscr{C}(\sigma)$ by the social equivalence relation $\sim_k \in \sim$. Further, let N be the total population of σ at a given time, and n the number of social equivalence relations (i.e. $N = |\mathscr{C}(\sigma)|$, $n = |\sim|$). Then the *degree of social differentiation* of σ with respect to $\sim$ at the given time equals

$$\delta(\sigma, \sim) = [2N(N-1)n^2]^{-1} \sum_{\sim_q \in \sim} \; \sum_{S_{pq} \in \mathscr{C}(\sigma)/\sim_q} \; \sum_{\sim_s \in \sim} \; \sum_{S_{rs} \in \mathscr{C}(\sigma)/\sim_s} | S_{pq} \triangle S_{rs}|.$$

This formula simplifies to

$$\delta(\sigma, \sim) = [n(N-1)]^{-1} \sum_k (n_k - 1),$$

where n_k is the number of cells produced by the kth partition, i.e. $n_k = |\mathscr{C}(\sigma)/\sim_k|$.

Minimal differentiation occurs when each partition produces a single class, namely $\mathscr{C}(\sigma)$ itself – i.e. when everyone is everything. In this case $\delta = 0$. At the other end of the spectrum there is just one individual in each social group, i.e. $|S_{ik}| = 1$ and $n_k = N$ for each pair (i, k). The corresponding value is $\delta = 1$. We take it that neither of these extreme values is realistic. The intermediate case is that in which all cells are uniformly populated, i.e. $|S_{ik}| = m$ and $n_k = N/m$, with $1 < m < N$ for every k. The corresponding degree of differentiation is

$$\delta(\sigma, \sim) = (N - m)/m(N - 1) \longrightarrow 1/m \quad \text{for } N \longrightarrow \infty.$$

A person who belongs to a social group is said to *participate* in it. Participation calls for differentiation: there is no occasion to participate if there are no different social groups. (However, participation is also nil if differentiation is maximal, since in the latter case no individual belongs to more than a single social group.) The individual participations aggregate to form a systemic property that may be called the *overall participation* in a society. This is made up of the overlaps of all the social groups, i.e. of $S_{pq} \cap S_{rs}$. More precisely, we make

DEFINITION 5.27 Let σ be a society whose composition is partitioned by the n equivalence relations contained in $\sim$. Then

(i) the *degree of participation of group* S_{pq} in σ is

$$\pi(S_{pq}, \sigma, \sim) = \sum_{\sim_s \in \sim} \sum_{S_{rs} \in \mathscr{C}(\sigma)/\sim_s} |S_{pq} \cap S_{rs}| \, / \, |S_{rs}|;$$

(ii) the *degree of marginality* (or *alienation*) of group S_{pq} in σ is the complement to unity of its degree of participation, i.e.

$$\mu(S_{pq}, \sigma, \sim) = 1 - \pi(S_{pq}, \sigma, \sim);$$

(iii) the degree of *overall participation* in σ is

$$\pi(\sigma, \sim) = F \sum_{\sim_q \in \sim} \sum_{S_{pq} \in \mathscr{C}(\sigma)/\sim_q} \sum_{\sim_s \in \sim} \sum_{S_{rs} \in \mathscr{C}(\sigma)/\sim_s} |S_{pq} \cap S_{rs}|,$$

where F is the normalization factor ensuring that the maximal value of π is 1;

(iv) the degree of *overall margination* (or *alienation*) in σ is

$$\mu(\sigma, \sim) = 1 - \pi(\sigma, \sim).$$

From the findings of social scientists we infer that all societies are differentiated, though none maximally, and that there is some measure of participation in all. That is, we propose

POSTULATE 5.9 For every society σ, $|\sim| > 1$, and $0 < \delta(\sigma, \sim) < 1$, $0 < \pi(\sigma, \sim) < 1$.

Every society is more or less cohesive: were it not, it would not be a system. Social cohesion is of course one of the emergent societal properties: it is a systemic property not possessed by any of the individuals composing a society. Cohesion can be spontaneous, forced, or a combination of the two. Spontaneous cohesion, the one that does not result from coercion, is undoubtedly the more important of the three. When strong it ensures the stability of the social fabric; when weak it calls for either structural change or the violent repression of any attempt to effect structural changes.

In Ch. 4 we assumed that cohesion derives from sharing (of things and services) and participation (in activities): recall Postulate 4.24. We shall presently discuss a measure of spontaneous cohesion based on that assumption (Bunge and García-Sucre, 1976). No participation at all, i.e. extreme segregation, results in instability, which ends up in fragmentation. Yet maximal participation has a similar dissolving effect, for a society of equals in every respect is a collection of competitors in no need of mutual assistance, i.e. in no need of belonging to the society. In other words, if every individual is either nothing or everything in a society, i.e. utterly marginal or ubiquitous, then cohesion is minimal. Maximal cohesion must then occur somewhere in between, when each individual shares several, though by no means all, cells with fellow members. In sum, cohesion must result from moderate participation. We proceed to formalize this intuitive idea.

The more members two social groups share the stronger they are bonded. However, these bonds operate only if the groups are different, in particular functionally complementary to one another. More precisely, the effective bonding between two social groups A and B is the numerosity of their intersection times that of their symmetric difference, i.e. $|A \cap B| \cdot |A \triangle B|$. A low degree of differentiation (i.e. a small $|A \triangle B|$ value) may be compensated for by a high degree of participation (i.e. a large overlap value $|A \cap B|$) and conversely. Certainly differentiation by itself does not contribute to cohesion, but the various partitions (structures) may intertwine in such a manner that cohesion results: i.e. cells situated in different columns of the social structure matrix may overlap partially. By adding up all the pairwise bonds among the social groups of a society we obtain its cohesion:

DEFINITION 5.28 Let σ be a society whose composition is partitioned by the n social equivalence relations contained in $\sim$. Then the degree of *spontaneous social cohesion* of σ with respect to $\sim$ equals

$$\kappa_s(\sigma, \sim) = [2N(N-1)n^2]^{-1} \sum_{\sim_q \in \sim} \sum_{S_{pq} \in \mathscr{C}(\sigma)/\sim_q} \sum_{\sim_s \in \sim} \sum_{S_{rs} \in \mathscr{C}(\sigma)/\sim_s} |S_{pq} \cap S_{rs}| \cdot |S_{pq} \triangle S_{rs}|,$$

where the *pqrs* term in the summation represents the strength of the bond between cells S_{pq} and S_{rs}, this strength being the product of the social participation by the differentiation.

Disjoint cells do not contribute to cohesion while overlapping cells do, provided they differ somewhat. If the nonempty overlaps all have unit measure, i.e. $|S_{pq} \cap S_{rs}| = 1$ for all pairs of cells except those along the same column of the social structure matrix, then cohesion reduces to differentiation (Definition 5.26). If on the other hand all the local differences are the same, and in particular $|S_{pq} \triangle S_{rs}| = 1$, then cohesion coincides with participation (Definition 5.27).

Minimal differentiation accompanies minimal cohesion. In fact, if $|S_{ik}| = N$ and $n_k = |\mathscr{C}(\sigma) / \sim_k| = 1$ for every k, then $S_{pq} = \mathscr{C}(\sigma)$ regardless of q, whence $S_{pq} \triangle S_{rs} = \varnothing$ in all cases, whence $\kappa_s(\sigma, \sim) = 0$. Likewise maximal differentiation ensues in a total lack of cohesion. Indeed, if $|S_{ik}| = 1$ and $n_k = |\mathscr{C}(\sigma) / \sim_k| = N$ for every k, the overlaps are nonvoid just in case the cells coincide, which implies that the symmetric difference is nil. In sum, in this case too $\kappa_s(\sigma, \sim) = 0$. Finally consider the intermediate case of homogeneity: $|S_{ik}| = m$ and $n_k = |\mathscr{C}(\sigma) / \sim_k| = N/m$ for every k, and $|S_{pq} \cap S_{rs}| = t$ unless the cells are in the same column. It turns out that in this case

$$\kappa_s(\sigma, \sim) = Nt(m-t)/(N-1)m^2 \cong t(m-t)/m^2.$$

The maximal value of spontaneous cohesion for this homogeneous case occurs for $t = m/2$ and is

$$\kappa_s(\sigma, \sim) = \tfrac{1}{4}N/(N-1) \cong \tfrac{1}{4}.$$

In short, the spontaneous cohesiveness of a society consists in a moderate participation of its members in the various social groups of the community. If the latter is hardly differentiated, there is not much to participate in, and competition overpowers cooperation, whence cohesiveness is low. And if the differentiation is rigid, as is the case with hierarchical societies (e.g. caste systems), there is not much participation either: in this case stability does not result from spontaneous cohesiveness but is imposed by force of arms or of ideology.

In contrast to spontaneous cohesion, forced cohesion is related to cultural or political repression, i.e. to the action of organized ideologies (e.g. churches and parties) or to police forces and the like. Though forced cohesiveness is an output variable, it can be roughly indicated by the corresponding input, namely the strength of the repressive forces relative to the total population. Calling $\kappa_f(\sigma)$ whatever indicator of the degree of forced cohesion we may choose, we can define the *total cohesiveness* of σ as $\kappa(\sigma) = \kappa_s(\sigma) + \kappa_f(\sigma)$. (In disregarding the set $\sim$ of social equivalence relations used to compute $\kappa_s(\sigma)$ we are tacitly assuming that all "significant" relations of that kind are being taken care of.) In principle, $\kappa(\sigma)$ is comprised between 0 and 2.

There are a number of definitions of the concept of a free society. Since most of them are ideological rather than scientific, there is little hope of achieving consensus. Thus while X will rightly count the freedom of the press among the necessary conditions for a free society, Y will counter, equally correctly, that such freedom is illusory as long as it can be exercised only by those with the economic freedom to run the press. We can avoid this debate by proposing an objective and even quantitative criterion of *political freedom*, namely this: A society is *politically free* iff its spontaneous cohesiveness is much greater than its forced cohesiveness.

If a society is not politically free, i.e. if its social structure is kept or modified by force of arms or ideology, then this force overpowers, balances, or eventually elicits a rebellion on the part of some social groups. However, the strength of the disruptive forces in a society is not just proportional to its numbers. Indeed, a small band of enthusiastic and well organized revolutionaries espousing a popular cause, or else acting amidst popular indifference, may defeat a large army of demoralized supporters of an unpopular dictatorship, just as a large mass of disorganized rebels will be no match for a small but highly trained repressive force. Consequently the politologist, unlike the sociologist, must study not only the indicators of objective cohesiveness but also those of "perceived" cohesiveness, or adherence (Sec. 3.3).

4.4. *Social Relations and Structures: Real?*

While holists acknowledge the reality of social relations, they reduce them to two types: the whole–whole (or system to system) relations – exemplified by international relations – and the whole– part causal relation, allegedly exemplified by group pressure. They are interested more in mythical wholes hovering above individuals – such as the Nation, the State, and the Spirit of the Times – than in interpersonal relations. The latter are on the other hand stressed by individualists: "Repeated interactions between particular persons are the very guts and marrow of social life" (Homans, 1974, p. 57). The systemist gives full assent to this view but he also admits intersystem bonds (e.g., Postulate 5.8(ii)). And he insists that social bonds, whether interpersonal or intersystemic, are real as long as they accompany changes in the relata (individuals or sociosystems). Consequently social structures are perfectly real, and moreover the main concern of the sociologist.

Social scientists rarely espouse the extremes of holism and individualism: most of them study social systems and their subsystems. And, whether they sympathize with holism or with individualism, they rarely deny the reality of social relations and, as a consequence, the reality of social structures. (Structuralist anthropology is an exception: it claims that social structures are in the head of the anthropologist. See Lévi-Strauss, 1963. For a devastating criticism see Harris, 1968.)

On the other hand the philosophers of social science who favor ontological and methodological individualism tend to deny the reality of social bonds, hence of social structures, perhaps because of their eagerness to deny the reality of the wholes extolled by the holists. Thus Winch (1958, pp. 131 ff.) has claimed that *all* "social relations between men exist only in and through their ideas", and "fall into the same logical category as do relations between ideas", so that social science is a chapter of epistemology. And Popper (1974, p. 14) has stated that "social relations belong, in many ways, to what I have more recently called 'the third world', or 'world 3', the world of theories, of books, of ideas, of problems". We saw in Ch. 4, Sec. 4.5, that there is no such thing, because a genuine world – i.e. a system – cannot be composed of both concrete things, such as books, and fictions, such as ideas in themselves (i.e. apart from ideating brains). Here we shall note only how convenient it would be for all oppressors to be able to persuade their subjects that their bondage belongs in the world of fairy tales and mathematical formulas, so they could not possibly suffer from it, let alone rebel against it.

Sociologists, who study social relations and social structures, and politicians, who attempt to control them, cannot take seriously the claim that social bonds and the social organizations they constitute are fictions. They regard them as just as real as physical or chemical bonds. Hence, when intent on changing them they resort to social action rather than incantation.

5. SOCIAL CHANGE

5.1. *The Social Structure Approach*

A social change is a change in some of the social properties of a community, such as a change in cultural level, or a redistribution of economic or political power. In principle it is possible to account for social change with the help of the concept of state function: the instantaneous state of a society would be the value of its state function at the given time, and the history of the society over a certain period would be the list of all the instantaneous values of its state function. This approach to social change should work for small sociosystems. But even a rough determination of the overall economic, cultural and political state of a modern society would require a staggering amount of information. And an account of its evolution would require an equally staggering number of equations of evolution (which in the simplest case would be rate equations for the state functions). We shall therefore adopt a more feasible method.

We shall assume that every social change consists in the redistribution of the composition of a society among the various cells of the social matrix $\|S_{ik}\|$. (For example, the emergence or the disappearance of a given kind of work, or of a given power relation, will ensue in the formation or extinction of the corresponding social cells.) Hence tracing the evolution of a society boils down to taking a periodic census – which is of course what the census bureaus do. This procedure will give us a precise, albeit coarse, description, though not an explanation, of social change. More precisely, we can make

DEFINITION 5.29 Let σ be a society with social structure matrix $\|S_{ik}\|$ and corresponding population density matrix

$$D(\sigma, t) = (1/N(\sigma, t)) \cdot \|N_{ik}(\sigma, t)\|.$$

The *net structural change* in σ over the time interval $[t_1, t_2]$ is

$$\Delta(\sigma; t_1, t_2) = D(\sigma, t_2) - D(\sigma, t_1).$$

Example At t_1, σ is divided into two social classes: masters (M) and slaves (S). At time $t_2 > t_1$, σ is leveled to a society of free men (F) with conservation of the total population. (I.e. either the revolution has been bloodless or the losses in lives have been compensated for by births or immigration.) The net class change (or change in class structure) is then

$$(1/F) \cdot \begin{Vmatrix} 0 \\ 0 \\ F \end{Vmatrix} - (1/F) \cdot \begin{Vmatrix} M \\ S \\ O \end{Vmatrix} = (1/F) \cdot \begin{Vmatrix} -M \\ -S \\ F \end{Vmatrix}, \quad F = M + S.$$

A negative (positive) sign indicates population loss (gain) in a cell. It goes without saying that, by virtue of the interrelations among the various social properties, a full account of this (imaginary) social revolution would require an investigation of all the major social cells.

Whatever causes, natural or social, may influence the distribution of the total population of a society among its cells, may be lumped into a social evolution operator. We assume that this operator satisfies

POSTULATE 5.10 Let $D(\sigma, t_1)$ and $D(\sigma, t_2)$ represent the social structure of a society σ at times t_1 and t_2 respectively. Then there exists an operator $E(\sigma; t_1, t_2)$ such that

$$D(\sigma, t_2) = E(\sigma; t_1, t_2) D(\sigma, t_1)$$

representing the mode of change of σ over the period $[t_1, t_2]$ and called the *social evolution operator* for that period.

This is a substantive assumption that may be found empirically wanting, particularly over a long period: the actual social evolution may well be nonlinear for all we know. However, we may assume that the postulate holds at least for short periods.

From the above postulate and Definition 5.29 it follows that the net structural change is

$$\Delta(\sigma; t_2, t_1) = (E(\sigma; t_1, t_2) - I)\, D(\sigma, t_1),$$

where I is a suitable unit matrix. Given the initial state $D(\sigma, t_1)$ and the mode of evolution $E(\sigma; t_1, t_2)$, the final structure, i.e. $D(\sigma, t_2)$, remains uniquely determined. This is the direct or forecast problem. The inverse or explanation problem is this: Given the values of the density matrix at two different times, e.g. now and one decade ago, find the mode of evolution, i.e. compute the evolution operator. Being an inverse problem, in general it lacks a unique solution and is much harder than the forecast problem. In fact, consider again the example studied a moment ago. The following evolution operators, among others, account for the same structural change:

$$E = \begin{Vmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 1 & 1 & 0 \end{Vmatrix} \quad \text{and} \quad E = \begin{Vmatrix} -1 & 0 & 1 \\ 0 & -1 & 1 \\ 1 & 1 & 1 \end{Vmatrix}.$$

It can be shown in general that the following holds (Bunge, 1974d):

THEOREM 5.1 Any given or prescribed social change in a society can be brought about in alternative ways, each represented by a different social evolution operator.

Further, since the populations along any given column of the social structure matrix must add up to the total population, we also have

THEOREM 5.2 The relative growth of any cell in a column of the social structure matrix occurs at the expense of the relative decay of some other cell(s) in the same column.

For example, if a column consists of just two cells and one of them grows exponentially, then the other must decay exponentially. Thus all growth has to be paid for by some decline.

Finally, note that social change is the likelier the more heterogeneous or varied a society is, and this if only because the more social cells there are, the more numerous the possible transitions among them. For the same reason, the more homogeneous a society the less changeable it is.

5.2. *History*

The main types of structural change of a society are listed in

DEFINITION 5.30 Let σ be a society with social matrix $\|S_{ik}\|$ and population matrix $\|N_{ik}\|$. Then

(i) σ remains *stationary* (or in a state of *static equilibrium*) in the kth respect (where $1 \leq k \leq n$) over the period $[0, \tau]$ iff $N_{ik}(t) =$ constant for all $1 \leq i \leq n$ and all $t \in [0, \tau]$;

(ii) σ is stable (or in a *state* of *dynamical equilibrium*) in the kth respect $(1 \leq k \leq n)$ over the period $[0, \tau]$ iff all the cells in the kth column of $\|S_{ik}\|$ expand or shrink at the same rate:

$$N_{ik}(t) = at \quad \text{with} \quad a \in \mathbb{R}^+ \quad \text{and} \quad a \neq 1 \quad \text{for all } 1 \leq i \leq m;$$

(iii) σ undergoes *cyclic changes* iff there is an evolution operator E and a natural number p such that $p = \tau$ and $E^p = I$;

(iv) σ undergoes an *irreversible* change iff σ changes in a noncyclic way;

(v) σ undergoes a *partial revolution* iff some cells S_{ik} change appreciably (in particular if they are either filled up or emptied);

(vi) σ undergoes a *total revolution* iff all the cells S_{ik} change appreciably.

Example A society consists initially of four social classes, 1, 2, 3, and 4, with populations N_1, N_2, N_3, and N_4 respectively. A revolution wipes out all but a fraction f_i of each class, and merges 1 with 2, and 3 with 4. A possible representation of the net change is this:

$$\begin{Vmatrix} f_1 & f_2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & f_3 & f_4 \\ 0 & 0 & 0 & 0 \end{Vmatrix} \begin{Vmatrix} N_1 \\ N_2 \\ N_3 \\ N_4 \end{Vmatrix} = \begin{Vmatrix} f_1N_1 + f_2N_2 \\ 0 \\ f_3N_3 + f_4N_4 \\ 0 \end{Vmatrix}$$

$$V = 1 - (1/N(0)) \textstyle\sum_i f_i N_i(0).$$

This index equals 0 for peaceful revolutions, $\frac{1}{10}$ for those that decimate each social stratum, and $\frac{1}{2}$ for those that take the lives of half the population in each group.

So far we have drawn no distinction between spontaneous and planned social change – hence between descriptive models and prescriptive ones. But all contemporary societies, even those with economic free enterprise, are subjected to planning to some degree or other, and it may be argued that they need more (and better) of the same if they are to survive. Take the rather obvious case of population growth. In the absence of any social or cultural constraints, and pretending that there are no environmental strictures either – in particular no interpopulational competition – the expected population size satisfies the rate law

$$\dot{N} = aN, \quad \text{with} \quad a = \text{birth rate} - \text{death rate}.$$

Here $a \in \mathbb{R}$ is a parameter summarizing biological, economic, cultural, and political characteristics, such as fertility, health level, food supply, and feeling of security. If $a > 0$, the population grows exponentially, if $a = 0$ it has zero growth, and if $a < 0$ it declines exponentially. Add now family planning on an individual basis, such as practised in all industrialized nations. Anything can happen depending on economic, cultural and political changes: the expected population size may increase, decrease, oscillate, remain stagnant, or neither.

Consider next family planning associated with the adoption of an explicit desideratum concerning the desirable family size. This norm will result in a total population goal G. A new law of population growth emerges, one that is controlled by the families themselves jointly with state organizations. In the simplest case this law takes the form

$$\dot{N} = b(G - N)\begin{cases} > 0 & \text{iff } N < G \text{ (underpopulation)} \\ = 0 & \text{iff } N = G \text{ (attained ideal)} \\ < 0 & \text{iff } N > G \text{ (overpopulation).} \end{cases}$$

Now population growth is under control. The total population grows exponentially (barring war, pestilence or famine) just in case the actual population is inferior to the goal value. It remains constant if the ideal has been attained. And it declines as soon as the goal is surpassed.

The above pattern of planned social change can obviously be generalized to an arbitrary social cell, such as a given occupational or educational group. As a matter of fact all modern nations do plan and control a number of activities, particularly agriculture and education: just think of the subsidies to the former and of the restrictions on entrance to medical schools.

Note the following points. First, mathematics can handle goals or norms; moreover, only with the help of mathematical models can social change be steered successfully. Second, knowledge and valuation can change the laws of social change. Such modifications consist, essentially, in the deliberate introduction of feedback terms of the form $G - X$, where X is the current value and G the corresponding goal value of a state variable. Third, controls are not restricted to temporal patterns but can be generalized to changes of all kinds. For example, if the spatial distribution of a social cell (e.g. farmers, craftsmen, or medical doctors) is to be controlled, the social planner will regard the density of each cell as a function of both spatial and temporal coordinates, and will set up diffusion equations containing feedback or control terms.

Whether spontaneous or planned, the social changes occurring in each social cell of a society form a time series or temporal sequence. And the whole bundle of such partial social histories, one per cell, constitutes the global social history of the given society. More precisely, we make

DEFINITION 5.31 Let σ be a society with instantaneous social structure (or density matrix) $D(\sigma, t)$. Then

(i) the *history* of the S_{ik} cell of σ throughout the period $[0, \tau]$ is the sequence

$$H_{ik}(\sigma; 0, \tau) = \langle D_{ik}(\sigma, t) \mid t \in [0, \tau] \rangle;$$

(ii) the *social* (or *anonymous*) *history* of σ over the period $[0, \tau]$ is the matrix $H = \|H_{ik}\|$ composed of all the social cell histories.

In other words, by collecting the values of the density of a cell over a period of time we obtain the history of that cell during that period. And doing the same for every cell in the social structure we come up with the social history of the society. (This may be dubbed the 'demographic view' of social history.) Obviously it will include the net effects of the contributions of the individual components of the society, though not their biographies. I.e. social history does not display the accomplishments of great men or great crooks: it is anonymous, therefore harder to reconstruct than biography. It attempts to describe, and eventually to explain, the evolution of entire cells, societies, and supersystems of such. It attacks big problems, in particular the coming into being and evolution of entire social cells, organizations, and institutions. For example, social history is concerned with the origin of agriculture and of capitalism, the origin of the state and of organized religion, the expansion of the world market and advanced technology.

A systemic approach to social history encourages studying the evolution of systems proper rather than artificial units such as many a kingdom resulting from war or political compromise. Such social systems may be circumscribed such as the Florentine banking system during the Renaissance, or vast like the Mediterranean basin during the same period (cf. Braudel, 1949). And, of course, a systemic approach will emphasize rather than ignore the interdependence among the various subsystems. For example it will study the evolution of the military–industrial complex, i.e. the symbiotic system composed of certain industries and the military. Or it will investigate the various facets of the Green Revolution, which started in the laboratory, changed the agrotechnics of India and elsewhere, and unexpectedly ended by worsening the lot of the poor peasants. So much for piecemeal social engineering, or the reform policy that ignores the interrelationships among the economic, the cultural, and the political subsystems of any society.

6. CONCLUDING REMARKS

We have sketched a systemic view of society, which may be regarded as a synthesis of two traditional conceptions: individualism (or atomism), and collectivism (or holism). In fact we regard society or any of its subsystems neither as an aggregate of individuals nor as a mystic totality hovering above its humble membership, but rather as a system composed of persons bonded by social relations. These ties generate social equivalence relations, such as those of belonging to the same economic, cultural, or political group. And these relations in turn partition the composition of a society, at a given time, into social cells or equivalence classes – such as the group of destitutes, or of illiterates, or of people deprived of political rights. Every such partition, or set of social cells, constitutes a social structure, or rather a primary social structure. A secondary social structure rides on the former, as when the set of families is classed into equivalent income groups. The totality of social structures of a society, both primary and secondary, constitutes its social fabric.

The systemic view of society should not be mistaken for the *organismic* view, a variant of holism according to which sociosystems are organisms. Surely the components of sociosystems are alive; besides, sociosystems are born, develop, and become extinct. However, they are not alive: if they were then all of the peculiar laws of the living, and in the first place those of genetics, would hold for them. But these do not, hence sociosystems are not alive. Nor are they dead: the categories of life and death do not apply to sociosystems. This is not to say that sociosystems are suprabiological, in particular spiritual entities, or that they are infrabiological, in particular chemical, systems. Sociosystems are concrete systems with peculiar properties and laws. They are concrete because they are composed of concrete things interacting in concrete ways and because they transform energy and effect changes in their environment. But they are as little physical or chemical systems as they are biosystems. This view of sociosystems may be christened *systemic materialism*, provided 'systemic' is not mistaken for 'holist' and 'materialism' is not construed in the narrow physicalist way.

We have introduced a method for both describing the structure of a society and keeping track of its multifarious changes, namely the social structure approach. The individualist is bound to dislike it because it achieves what he holds to be impossible, namely describing the state, and the changes of state, of a society as a whole. But he cannot in good faith dismiss it as being "merely a holistic confusion", for it is based on a conception of society which, being systemic, is neither holist nor individualist. Besides, the method is mathematically transparent, so it cannot be branded as confused – in particular because it does not involve the individualistic confusion of a sociosystem with its composition.

Our structural approach to social change ensues in a view of social (or anonymous) history which is as precise as it is shallow: it only recounts how the social cells are populated or depleted. But at least this view of social history is superior to both the great hero myth and the great whole myth. It is moreover consistent with contemporary social history and social demography. And far from ignoring the individual it encompasses every individual contribution with a social impact – i.e. that influences the social behavior of other individuals. Finally, nothing prevents the social scientist from supplementing our descriptive account of social structure and social change with mechanismic models capable of explaining the current structure and its development.

Besides a general framework for the study of social systems, we have introduced definite measures of certain emergent social properties, such as social differentiation, participation, and cohesion. And we have submitted that every society, however primitive, is made up of four main subsystems: the kinship, economic, cultural, and political ones. These are not independent spheres but interdependent subsystems. Therefore the neglect or ill health of any of them will sooner or later lead to a deterioration of the others. And, because of the feedback links among the various subsystems, favoring either at the expense of the others will eventually result in a decay of the very favored subsystem.

Thus *economism* ensues in the wilting of culture, which may include a technology indispensable to keeping the economy up to date; and it will tend to leave politics in the hands of "those who know best", thus fostering a concentration of power that will eventually stifle both the economy and the culture. Likewise *politicism*, or the exaggeration of the role of politics, may lead to a disruption of the very economy and culture that politics is supposed to serve, to the point that not much worth being managed may remain. Analogously *culturalism* would end up by draining the economy and drying up the political life, which would backfire on culture itself by depriving it of sustenance and political influence. All this, which is obvious from our systemic point of view, entails a moral that has yet to be learned by most political, economic, and cultural leaders, to wit: In a healthy society there is a balance among its main subsystems, even during periods of vigorous growth. In particular, there is no unilateral (e.g. exclusively economic) development – nor is development the same as growth.

Although we have proposed a handful of substantive hypotheses on the nature of sociosystems, ours is mainly a framework for the study of things of that kind, in particular for the building of theories in social science. (See Bunge (1974d) for the difference between framework and theory.) And surely no framework, however suggestive, is a substitute for a full fledged testable theory. Yet while some frameworks block, others facilitate the conception of theories. Ours discharges its heuristic function in the following ways:

(i) It invites studying social systems rather than either outstanding individuals or amorphous sets of persons or totalities way above individuals. In this way it stimulates the search for societal properties and laws.

(ii) It directs one to identifying the coordinates of every social system – its composition, environment (natural and artificial), and structure. In this way neither the people nor the environment nor the social reslationships can be lost sight of.

(iii) It stimulates the analysis of extremely large and complex systems into more manageable subsystems with their own inner dynamics as well as their mutual actions. Without such analyses hardly any scientific study of society is possible.

(iv) It suggests spotting and characterizing the three chief artificial subsystems of every society – the economy, the culture, and the polity. Thus it helps avoiding the pitfalls of economism, culturalism, and politicism.

(v) It guides the search for relationships among the components of each subsystem as well as among the subsystems of every system. Hence it avoids the fallacies of the lone wolf and the isolated organization.

(vi) It clarifies and simplifies the study of the so-called cross relations, or relations among relations, by constructing them as links among system components or subsystems – with considerable gain in clarity and economy.

(vii) It makes room for cultural, in particular ideological, influences in every walk of social life. But it avoids the reification of cultural activities and is itself ideology-free.

(viii) Although it is ideology-free, our framework does suggest certain social policies, such as that of combining equity with participation in the interests of social cohesion and therefore stability.

This concludes our study of sociosystems and at the same time that of all other genera of concrete system. We return now to the general concept of a concrete system, with the advantage of our realization of the diversity of kinds of system, hence of change.

CHAPTER 6

A SYSTEMIC WORLD VIEW

In this, the last chapter, we shall generalize some of the results of the previous chapters. We shall thus come up with the core of a systemic world view. This *Weltanschauung* is a continuation of certain philosophical traditions and is also in tune with contemporary science. But it should not be mistaken for the popular "systems philosophy," a new version of holism according to which every thing is a system (false) and the patterns of being and becoming are basically the same at all levels (false). Our systemist philosophy is neither holistic nor atomistic: it acknowledges the variety of properties, kinds and patterns found in the world and, by using certain elementary formal tools, it avoids the obscurities of traditional philosophy. Ours is, in short, a kind of scientific ontology. (Cf. Vol. 3, Introduction.) Let us review some of its assumptions.

6.1. *A World of Systems*

An assumption of this book and its companion, *The Furniture of the World* (Vol. 3), is that there are no stray things: that every thing interacts with other things, so that all things cohere forming systems. That is, we propose

POSTULATE 6.1 Every concrete thing is either a system or a component of one.

Another assumption is that systems come in Chinese boxes or nested systems. Hence, given any system but the world, one may expect to find that it is part of some supersystem. (The converse is false: not every system component is itself a system. There seem to be basic or simple things, such as electrons and photons.) Thus we lay down

POSTULATE 6.2 Every system except the universe is a subsystem of some other system.

Finally, the universe itself is neither a mere thing nor a loose aggregate of things but a system of systems:

POSTULATE 6.3 The universe is a system – namely the system such that every other thing is a component of it.

From the previous assumptions, together with some of our axioms on things and change (Vol. 3), a number of further sweeping generalizations follow. One such theorem is that every system is engaged in some process or other. Second: Every change in any system in lawful. Third: Because every subsystem acts upon or is acted on by other subsystems, it exists somehow *per aliud* (by another) rather than autonomously. Fourth: The universe as a whole exists *per se* (by itself) – and moreover it is the only absolute (autonomous) existent. Fifth: The universe has neither beginning nor end in time: it endures eternally although no part of its does.

6.2. *System Genera*

Philosophers are not equipped to ascertain what kinds of system there are or could be in the world: they can only take note of the variety of systems that science discovers, and help categorize them. Current science, pure and applied, seems to accept the existence of five system genera. (Recall that a kind, such as a species or a genus, is said to exist just in case it is not empty. Strictly speaking only the members of a kind can be said to exist really. Cf. Vol. 3, Ch. 3, Sec. 4.3.) The system genera characterized in the previous chapters are:

S_1 = *Physical*
Microphysical, e.g. atoms and small scale fields
Mesophysical, e.g. bodies and large scale fields
Megaphysical, e.g. galaxies
S_2 = *Chemical*
Microchemical e.g. chemical reactors involving only monomers
Mesochemical, e.g. chemical reactors involving polymers
Megachemical, e.g. compost piles
S_3 = *Biological*
Microbiological, e.g. single cells
Mesobiological, e.g. individual multicellular organisms
Megabiological, e.g. ecosystems
S_4 = *Social*
Microsocial, e.g. families
Mesosocial, e.g. villages and concerns
Megasocial, e.g. large cities and nations
S_5 = *Technical*
Microtechnical, e.g. this book

Mesotechnical, e.g. farms and industrial plants

Megatechnical, e.g. a lumber-paper manufacturing-and-commercialization system.

Remark 1 We have not lumped physical and chemical systems into a single category for the following reasons. Firstly, not all physical entities are systems. Secondly, chemical systems are never at rest: if all chemical reactions in a system come to a halt, the system becomes a physical one. Thirdly, chemosystems have hardly any inertia, as shown by their quickness to respond to external inputs and to the cessation of the latter. (Mathematically: the equations of chemical kinetics are of the first order in the rates, whereas most physical equations of evolution are of the second order or higher.) Fourthly, controllers, such as catalyzers, play an important role in chemosystems, for even the H^+ and OH^- ions are catalyzers. On the other hand the only physical systems endowed with control mechanisms seem to be artificial. *Remark 2* We might have distinguished a system genus between biosystems and sociosystems, namely psychosystems. We have refrained from doing so from fear of encouraging the myth of disembodied minds. In our view psychosystems belong in the biosystem genus because they are animals endowed with a highly evolved nervous system. *Remark 3* We have not grouped technical systems, i.e. artifacts, together with any others, because they have ontological traits that set them apart: they are a product of human work and, as such, they bear the stamp of human intelligence and purposiveness as well as of social organization.

We compress the above remarks into

POSTULATE 6.4 At the present stage of the evolution of the universe there are five system genera: physical, chemical, biological, social, and technical. I.e. the family of system genera is $\mathcal{S} = \{S_1, S_2, S_3, S_4, S_5\}$.

The cautious clause about the present is a reminder that the higher kinds have not always existed (so far as we know), and may either become extinct or give rise to further system genera. Thus the exploration of outer space might reveal the existence of further system genera, and a nuclear holocaust would wipe out S_3, S_4 and S_5 on our planet.

6.3. *Novelty Sources*

Even assuming that the universe has always contained things of certain kinds, such as protons, there is little doubt that things of new kinds come into being every so often. I.e. at least part of the marvelous variety of the world is a product of change. Now, there are two main kinds of change: quantitative (e.g. motion) and qualitative (e.g. chemical combination). In the former case all of the axes of the state space of the thing are used; in the latter some new axes pop up and others drop out as change proceeds. However, even quantitative change may end up in qualitative newness, as in the case of high speed collisions of atomic nuclei, the reshuffling of the components of a system, and growth by accretion or multiplication. Let us take a look at reshuffling and accretion, which would seem to be unlikely emergence mechanisms.

A *reorganization* or *restructuring* is of course a change in the structure of a system, i.e. a change modifying some of the links among the system components even though it may not alter appreciably the intrinsic properties of the said components. Two simple examples are isomeric transitions (in an atomic nucleus or in a molecule) and institutional reorganizations without changes in personnel. In a closed system, repeated shuffling ends up in a state of equilibrium and disorder where no novelty can occur. On the other hand in an open system certain reorganizations accompanied by qualitative novelty and increase of order can occur. *Examples* The cooling of a liquid until it solidifies forming a crystal; the cooling of a metallic ring until it becomes superconducting; the magnetization of an iron bar; and the formation of patterns in fluids and in chemical systems. (Cf. Turing, 1952; Glansdorff and Prigogine, 1971; Haken, 1975.)

Another mechanism that may ensue in qualitative novelty is accretion or assembly. Thus stars and planets are believed to have formed by gravitational accretion of cosmic gas and dust. However, novelty is more likely to emerge along processes where heterogeneous things come together to form systems. This may occur even if "mistakes" – i.e. departures from the common pattern – are possible, that jeopardize the completion of the assembly process. Now, in principle assembly can proceed either in one stage – whether instantly or gradually – or in several. For example, a one stage assembly process may consist in the addition of one component per unit time; and a multiple stage assembly process may start with the formation of decamers out of monomers, proceeding with the formation of hectomers out of decamers, and so on. If there is a nonvanishing probability that each item (whether a monomer, a decamer, etc.) joins another in the "wrong" way – e.g. forming an unstable system – then the direct or one stage process is unlikely to yield systems of high complexity. As the number of intermediate stages in the synthesis of a complex system increases, the probability of departure from normality decreases and so the probability of success increases. (For the Crane-Rosen mathematical model of multiple stage assembly, see Rosen, 1970.) Since this result is independent of the nature of the units that get assembled, as well as of the bonds among them, we may assume that all complex systems, whether living or nonliving, have assembled by stages. This general principle of system architectonics deserves being incorporated into our ontology:

POSTULATE 6.5 The more complex a system, the more numerous the stages in the process of its assembly.

The dual of this hypothesis concerns the breakdown or dismantling of systems: it states that complex systems can decay in different ways, each way corresponding to the weakening of one set of bonds among the components. (Even comparatively simple things such as mu mesons can have multiple disintegration schemata.) In short we lay down

POSTULATE 6.6 The more complex a system, the more numerous its possible breakdown modes.

6.4. *Emergence*

Our next assumption is that every concrete system has assembled from, or with the help of, things in the same or lower order genera. In order to formulate this assumption with some precision we need

DEFINITION 6.1 Let $S = \{S_1, S_2, S_3, S_4, S_5\}$ be the family of system genera. Then S_i *precedes* S_j, or $S_i < S_j$ for short, iff things of genus S_i take part (either as components or as agents) in the assembly of every system in S_j.

Our hypothesis is

POSTULATE 6.7 For any $x \in S_j$ there is at least one $y \in S_i$, where $S_i < S_j$, such that y has taken part in the assembly of x.

Thus far we have only asserted that every system has been preceded by pre-existing things – which is the reciprocal of the holistic thesis that the whole precedes its parts. But we have not said exactly what precedes what. This we do now:

POSTULATE 6.8 The precedence relations among the system genera are

$$S_1 < S_2 < S_3 < S_4, S_5.$$

That is, the family of system genera is a tree rooted in S_1 and branching out at S_3 into S_4 and S_5: see Figure 6.1.

The precedence relation $<$ is a partial order relation. Hence the ordered $\mathscr{S} = \langle S, < \rangle$ is a partially ordered set. It may be called the *system structure of the world.* This is often mistaken for a hierarchy. Yet the difference between the two concepts is clear: whereas $\mathscr{S} = \langle S, < \rangle$ is the set of system genera ordered by the precedence (or emergence) relation, a hierarchy is a set of things (or of sets of things) ordered by a dominance relation. Besides, the thesis that the world has a hierarchical structure is usually a form of supernaturalism, and is not countenanced by science. On the other hand the thesis that the concrete systems in the world are distributed among S, and that higher genus systems have emerged from lower genera ones, is a component of a naturalist, systemist, pluralist, and dynamicist ontology.

6.5. *Systemism Supersedes Atomism and Holism*

The eight postulates proposed in the preceding subsections match with the assumptions made in the preceding chapters and crown our version of systemism. This ontology must not be mistaken for holism, or the romantic extolling of wholeness and emergence criticized in Ch. 1, Sec. 4.2. For, while holism claims that wholes and emergents must be accepted with "reverence" (Goethe) or "natural piety" (Alexander), systemism encourages attempts to analyze systems into their composition, environment, and structure, as well as to disclose the mechanisms of their formation and breakdown.

The demand for explanations of the way of systems should not be mis-

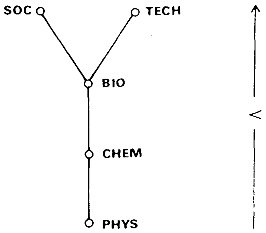


Fig. 6.1. The precedence relations among system genera.

taken for naive reductionism, i.e. the belief that all it takes to account for a whole is a knowledge of its parts. This belief is mistaken because systems possess properties absent from their components, hence not to be discovered by studying the latter in isolation from one another. Analysis is necessary but insufficient to account for the formation and persistence of systems: we must also study systems on their own level and interaction with things of different genera if we wish to understand them. This is what the theories of systems of various kinds – physical, chemical, biological, social, and technical – are all about.

The philosophy that combines an acknowledgment of emergence with the thesis that emergence is explainable and predictable within bounds may be called *rational emergentism.* This philosophy, foreshadowed by Sellars (1922), supersedes both atomism (though not its allegiance to science) and holism (though not its insistence on emergence), and it incorporates a critical realist theory of knowledge – of which more in the next two volumes of this *Treatise.*

6.6. *Synopsis*

It is closing time. We close by characterizing the ontology presented in this volume and its companion, *The Furniture of the World*, in terms of a few traditional *isms.* (A single *ism* is bad because it is one-sides and rigid. Many-sidedness and flexibility can be achieved only by a system of matching *isms.*)

Our ontology endorses

(i) *naturalism* or *materialism*, for it countenances only material existents and discounts autonomous ideas, ghosts, and the like; but *not physicalism* (or *mechanism*), as this denies that all things are physical entities;

(ii) *systemism*, for it holds that every thing is either a system or a component of one; but *not holism*, as it rejects the myths that the whole is incomprehensible, prior to its components, and superior to them;

(iii) *pluralism* as regards the variety of things and processes, hence the plurality of kinds of thing and laws; and also *monism* as regards the substance that possesses properties and undergoes change (namely matter) as well as the number of worlds (just one);

(iv) *emergentism* with regard to novelty, for it holds that, while some bulk properties of systems are resultant, others are emergent; but *not irrationalism* with regard to the possibility of explaining and predicting emergence;

(v) *dynamicism*, for it assumes that every thing is in flux in some respect or other; but *not dialectics*, for it rejects the tenets that every thing is a unity of opposites, and that every change consists in or is caused by some strife or ontic contradiction;

(vi) *evolutionism* with regard to the formation of systems of new kinds, for it maintains that new systems pop up all the time and are selected by their environment; but *neither gradualism nor saltationism*, for it recognizes both smooth changes and leaps;

(vii) *determinism* with regard to events and processes, by holding all of them to be lawful and none of them to come out of the blue or disappear without leaving traces; but *not causalism*, for it recognizes randomness and goal striving as types of process alongside causal ones;

(viii) *biosystemism* with regard to life, for it regards organisms as material systems that, though composed of chemosystems, have properties not found on other levels; but *neither vitalism nor machinism nor mechanism*;

(ix) *psychosystemism* with regard to mind, for it holds that mental functions are emergent activities (processes) of complex neural systems; but *neither eliminative nor reductive materialism*, for it affirms that the mental, though explainable with the help of physical, chemical, biological, and social premises, is emergent;

(x) *sociosystemism* with regard to society, for it claims that society is a system composed of subsystems (economy, culture, polity, etc.), and possessing properties (such as stratification and political stability) that no individual has; hence *neither individualism nor collectivism*, *neither idealism nor vulgar materialism*.

The reader accustomed to dwell in a single *ism* or in none is likely to throw up his hands in despair at the multiplicity of *isms* embraced by our ontology. Let this be said in defense of such multiplicity. First, it is possible to synthesize a variety of philosophical *isms* provided they are not mutually inconsistent – i.e. provided the result is a coherent conceptual system rather than an eclectic bag. (We have tried to secure consistency by adopting the axiomatic format.) Second, it is necessary to adopt (and elaborate) a number of philosophical *isms* to account for the variety and mutability of reality – provided the various theses harmonize with science. Third, tradition can be avoided only at the risk of unfairness and ignorance: rather than dismiss our philosophical legacy altogether, we should try and enrich it.

APPENDIX A

SYSTEM MODELS

This Appendix is a review and philosophical analysis of some simple yet serviceable models of concrete systems regardless of the nature of their components, hence of the kind of bonds among them. The aims of this presentation are (*a*) to clarify some of the key concepts employed in the text, (*b*) to supply a stock of examples, and (*c*) to discuss some problems in the as yet immature philosophy of systems science. For details see Ashby (1956), Athans and Falb (1966), Dorf (1974), Harrison (1965), Klir (1969), Klir and Valach (1967), Lange (1965), Padulo and Arbib (1974), and Zadeh and Desoer (1963).

1. INPUT-OUTPUT MODELS

1.1. *The Black Box*

The simplest (i.e. poorest) model of a thing one can think of is a black box that does not interact with its environment. ('Black' because it has no structure or we do not care to disclose it.) This, the bare black box model, ignores both the internal structure of the thing modeled and its actions upon other things as well as the actions its environment exerts on it. Despite its poverty, such a model can be endowed with a number of attributes assumed to represent properties of the real thing. The classical particle in free motion and the totally isolated thermostatic system are cases in point.

Next in order of complexity comes the black box proper, fitted with terminals connecting it to its environment – which is in turn treated globally or as composed of other black boxes. The simplest black box has just one input terminal, representing either the action (in a given respect) of the environment on the thing, or the thing's reaction (in the same or a different respect) upon the environment. The former model represents a passive thing, the latter an active one, and both are called *monopoles*. Since in reality there are neither totally passive nor fully active systems, the monopole model is adequate only in extreme cases and when a single property is considered.

A more realistic model of a thing is of course the black box fitted with both input and output terminals, i.e. sensitive to (a part of) its environment and in turn reacting on (some features of) the latter. We call this the *complete black box*, or *black box* for short. The simplest complete black box is sensitive to stimuli of just one kind and accepts just one stimulus at a time – i.e. has a single input terminal; and it also has a single output terminal, i.e. acts just in one fashion on its environment and performs one action at a time. This, the simplest black box, is called a *dipole*. Any more complex black box is called a *multipole*: it is equipped with a certain number of input terminals of the same or different kinds. See Figure A1. We shall concentrate our attention on the dipole, which is the building block of every other input-output model of a system.

From a strictly formal point of view a dipole is just a double entry table of items (e.g. numbers), or set of ordered pairs $\langle u, v\rangle$, where $u \in U$ is an input value and $v \in V$ the corresponding output value. In other words, a black box of this kind boils down to a binary relation f (perhaps a function) with graph or extension $\mathscr{E}(f) \subset U \times V$, where U is the collection of possible inputs and V that of possible outputs. (This is what Mesarović (1968) calls a *general system*. But it is a model of a system and moreover not completely general, for it neglects the internal structure of the system.)

The simplest of all such model systems, i.e. the dipole, has inputs and outputs of the yes or no type: i.e. the terminals are either on (value 1) or off (value 0). That is, $U = V = \mathbb{2} = \{0, 1\}$, and $f\colon \mathbb{2} \to \mathbb{2}$ is the *transfer function*. *Example 1* The conformist box: $f(0) = 0, f(1) = 1$. *Example 2* The nonconformist box: $f(0) = 1, f(1) = 0$. *Example 3* The capricious (yet not chaotic) or stochastic box: $Pr(0 \mid 0) = p > 0$, $Pr(1 \mid 0) = 1 - p$, $Pr(0 \mid 1) = q > 0$, $Pr(1 \mid 1) = 1 - q$, where p is the probability of response 0 to stimulus 0, and similarly for q. (Note the difference between the stochastic box, characterized by constant values of p and q, hence statistically reliable, and a totally chaotic box, i.e. one with no transfer function.) All three boxes are coarse models of a specialist; the third models the creative type.

Other black boxes accept continuous streams of stimuli and deliver

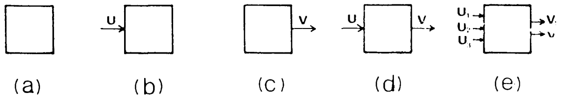


Fig. A1. (*a*) The bare black box. (*b*) The passive monopole, or system without outputs. (*c*) The spontaneous monopole, or system without inputs. (*d*) The dipole, or simplest complete black box. (*e*) The reactive multipole.

continuous streams of outputs. That is, their transfer functions are smooth functions of time. *Example 1* Differentiator. The state of the system at time $t \in T$ is the ordered pair $\langle u(t), \dot{u}(t)\rangle$, where $u : T \to \mathbb{R}$ and $\dot{u}$ is the rate of change of u. The state space of the system is $S = U \times V$, where U is the image of u and V that of $v = \dot{u}$. *Example 2* Delay line. Here $v(t) = u(t - \delta)$ with $\delta > 0$. I.e. the current output equals the input at the earlier time $t - \delta$. The state space is $S = U \times \{u(t - \delta) \mid t \in T\}$.

In the last two examples the input and output values were parametrized by time – i.e. U and V were construed as the images of time-dependent functions. In general the output value $v(t)$ at time t will depend not only on the current value of the input (as in Example 1) but also on some or even all of the inputs accepted earlier by the system (as in Example 2). I.e. in general the system will have *memory* (or be *hereditary*). A fairly general input–output relation covering all such cases for the dipole, and easily generalizable to the multipole, is

$$v(t) = \int_{-\infty}^{t} d\tau\, M(t, \tau) F(u(\tau)),$$

where M summarizes the structure of the box (Bunge, 1963a). The domain of integration, which stops at t, has been chosen to conform with the antecedence principle (Postulate 6.12 in Vol. 3, Ch. 6, Sec. 5.3). If M is a delta function, then the box is memoriless. If F is a linear function then the above represents a linear dipole. In a great many cases M does not depend on t but only on the past (if at all), and $F(u(\tau)) = u(t - \delta)$, where δ is a time lag. In such cases the integral is the convolution of the two functions, and the mathematics is greatly simplified.

1.2. *Connecting Black Boxes*

Strictly speaking a black box is a suitable model not of a system but rather of a system component. In many cases a system can be adequately modeled as an assembly of black boxes connected in a suitable way. The basic modes of coupling are as follows.

(i) *Series coupling*: see Figure A2a. If the two component boxes are characterized by the transfer functions $f : U \to V$ and $g : V \to W$, then the resulting box is characterized by the composition $h = f \circ g$ with values $w = g(f(u))$. Note that, in general, function composition is not commutative; thus g might filter out all stimuli except those in a certain set, while

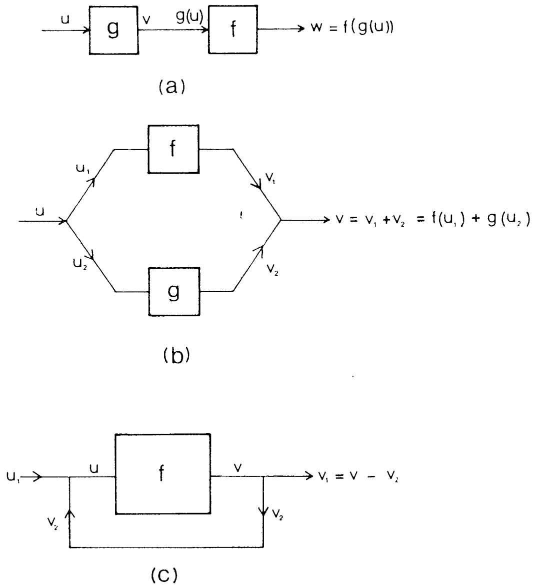


Fig. A2. (*a*) Series combination of two dipoles. Example: cloth manufacturer *g* supplies clothing manufacturer *f*. (*b*) Parallel connection of two dipoles. Example: two branches of a cloth manufacturing concern. (*c*) Feedback coupling (or self interconnection). Example: blood pressure control system.

f could just delay the responses. Example: a (living or industrial) plant *f* that manufactures materials produced by another plant *g*.

(ii) *Parallel coupling*: see Figure A2b. The input is shared by the two components and the output is the sum of the partial outputs. I.e. $u = u_1 + u_2$ and $v = f(u_1) + g(u_2)$. Example: two industrial plants that consume and produce similar goods at the same time.

(iii) *Feedback coupling*: see Figure A2c. The system is coupled to itself in such a way that part of its output is fed back as an input. That is, $v = f(u) = v_1 + v_2$ and $u = u_1 + v_2$, where now u_1 is the external input (stimulus) and v_2 the net output (response) whereas v_2 is the correcting feedback.

By combining tandem, parallel and feedback couplings one can synthesize any complex system of the input–output type. An assemblage of interconnected black boxes is not a black box itself but has emergent properties. One way of representing the resulting system is by means of the list of all the transfer functions characterizing its components when inter-connected, i.e.

$$\mathbb{F} = \langle f_1, f_2, \dots, f_n \rangle, \quad \text{where} \quad f_i\colon U_i \longrightarrow V_i.$$

While some of these functions represent external terminals (or *exogenous* variables), others represent internal terminals (or *endogenous* variables).

Input–output models are used in a number of fields, from thermodynamics and electrical engineering (the two historical sources) to biology, psychology, and social science. (Most of what goes under the name of *systems analysis* consists in building such models.) Their success has fostered the beliefs that (*a*) every system can be fully modeled by an input–output model, and (*b*) the world as a whole is a supersystem composed of multipoles. But these are illusions. (The technologist's sweet dream is the metaphysician's nightmare.) Firstly, there are plenty of systems without any recognizable inputs and outputs: just think of a radio wave. Secondly, in many cases we need to know the spatial locations of the parts of a system: time is not enough. (Think again of a radio wave, or of a brain, or of a city.) An input-output analysis, though practicable and useful in many areas – particularly in technology – is much too poor and is bound to be eventually supplemented or even superseded by a deeper and more detailed model. Thus thermodynamics is supplemented by statistical mechanics, electric network theory by electrodynamics, stimulus-response psychology by psychobiology, and so on. (For an evaluation of black boxes and a criticism of blackboxist philosophy see Bunge (1964).)

1.3. *Control System*

A control system is composed of two subsystems: an ordinary or controlled system and a controlling system, i.e. one that controls or regulates the output of the former. The controlling system acts by absorbing part of the total output and feeding it back into the system in such a way that the effective input itself is modified, and the net output differs from that of the uncontrolled system. The controlling system has two main components: the *imbalance* (or *error*) *detector* and the *response mechanism.* The latter is driven by the former in proportion to the deviation of the current state of the controlled system from the set point or ideal state or "goal". In other words, the detector is activated (has an output) just in case the controlled system is in a state other than the set point. See Figure A3.

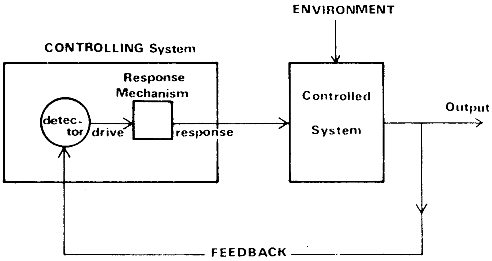


Fig. A3. Feedback control system. The detector compares the output (actual state) of the controlled system and, if it differs from the set point, drives the response or corrective device, which acts on the controlled system to correct the imbalance.

The study of specific control systems belongs to the special sciences, such as mechanical and electrical engineering, biochemistry, physiology, economics, and politology. And the scientific study of the features common to all control systems, regardless of their composition, is of course cybernetics (Wiener, 1948). Cybernetics was born in advanced engineering but, since it is stuff-free (i.e. independent of the kind of material), it assumes no special physical, chemical, biological, or social laws, so it covers all levels. It is therefore a chapter of the general theory of systems as well as of scientific ontology.

The supersystem composed by the controlled and the controlling systems can be described as follows. The total state function of the supersystem is

$$\mathbb{F} = \langle F_1, F_2, \dots, F_p, F_{p+1}, F_{p+2}, \dots, F_n \rangle : T \longrightarrow \mathbb{R}^n,$$

where the first p coordinates are the so-called *state variables* and the remaining $n - p$ components the *control variables*. These components are subjected to three sets of conditions:

(i) *constraints* on the state variables (e.g. their values must lie within a certain region of the state space);

(ii) *control* or *steering conditions* specifying the manner in which the control variables affect the state variables, e.g.

$$\dot{F}_i = g_i(\mathbb{F}, t) \quad \text{for } 1 \leq i \leq p;$$

(iii) *optimization conditions* on certain state variables (e.g. minimization of energy consumption, or maximization of the rate of production of some substance).

A good way to get a feel for cybernetics is to study a simple control system such as an integrator (e.g. area calculator) with feedback control. In the absence of a controller and if the system is initially inert, the input-output equation is

$$\dot{v} = u, \quad \text{with} \quad v(t) = 0 \quad \text{for } t \leq 0.$$

This entails

$$v(t) = \int_0^t d\tau \, u(\tau),$$

so that the system integrates the input. In particular, a constant input

$$u(t) = H(t) = \begin{cases} a \in \mathbb{R}^+ & \text{for } t \geq 0 \\ 0 & \text{for } t < 0 \end{cases}$$

is converted into $v(t) = at$. Obviously this output, unless checked, gets eventually out of hand. This can be prevented by coupling the integrator to a controller that diverts part of the output and sends it back to the input terminal, so that a closed loop results: see Figure A4.

In the simplest case the output fraction $kv(t)$, where $|k| < 1$, is fed back diminished but without distortion, so that the input–output equation reads now

$$v = u + kv, \quad \text{with} \quad -1 < k < 1.$$

If $k > 0$, the feedback is *positive*, i.e. it enhances the effect of the input, so that the system may become uncontrollable. The feedback is *negative*, and results in a stable system, if it checks the input, i.e. $k < 0$. The solutions for the case of the constant input $u = H(t)$ applied for $t > 0$ are

(i) for $k > 0$ (*positive feedback*)

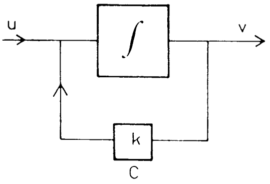


Fig. A4. Integrator with feedback control. The effect of the control subsystem C is to convert a constant input into either an exponentially growing response (for $k > 0$) or an exponentially decaying one (for $k < 0$).

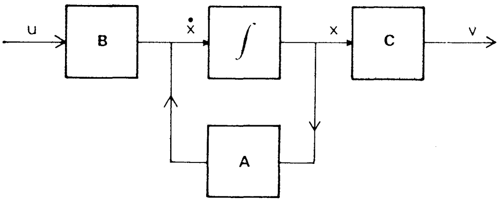


Fig. A5. Linear control system.

$$v(t) = \frac{a}{k}(\exp(kt) - 1) \quad \text{exponential growth}$$

(ii) for $k < 0$ (*negative feedback*)

$$v(t) = \frac{a}{|k|}(1 - \exp(-|k|t)) \quad \text{decelerated growth approaching the target value } a/|k|$$

Note that both the controlled and the controlling subsystems have been treated as black boxes: their mechanisms have not been specified. Nor could they without roping in definite law statements (physical, biological, or what have you).

The above example illustrates the black box model of a control system. In most cases such a model is insufficient and a grey box model is required, i.e. one involving not only an input vector u and an output vector v but also an (internal) state vector x. The simplest example of such a model of a control system is, of course, the linear system

$$\dot{x} = Ax + Bu, \qquad v = Cx,$$

where the state matrix A, the input matrix B, and the output matrix C are constant. The corresponding block diagram is shown in Figure A5.

1.4. *Stability and Breakdown*

As we saw in Sec. 1.3, feedback can be positive (enhancing) or negative (checking). If negative, feedback may correct misalignments. If it does, the corresponding system is said to be *self-stabilizing*, for in such case all the components of its state vector remain bounded all the time. See Figure A6.

Equilibrium can of course be static or dynamic. In the former case nothing changes, so that the representative point in the state space remains

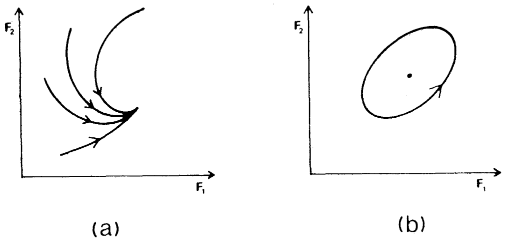


Fig. A6. Two kinds of stable orbits in state space. (a) Spiral approach to static equilibrium. (b) Stationary orbit.

fixed. (But of course no system remains forever in a state of static equilibrium for, even if it is internally stable, every system is subjected to disruptive external influences.) In the case of dynamic equilibrium the trajectories remain within a certain limited region of the state space. The departures from equilibrium may be regular (e.g. sinusoidal oscillations about a point) or irregular – in which case they are called *fluctuations* or *stochastic oscillations*. A system may oscillate around an equilibrium position, whether regularly or irregularly, either spontaneously (self-excited oscillations) or under the action of its environment (forced oscillations). If the oscillation amplitude increases in the course of time instead of remaining bounded or fading out (damping), the system is unstable and eventually breaks down.

If the ongoing existence of a system depends critically on its keeping its properties within certain bounds, then it will have more chances of "survival" if it is self-stabilizing than in any other case. But of course even a system of this kind may be subjected to such strong external disturbances that they drive it permanently far from equilibrium and eventually into breakdown. If the parameters of the controlling components (e.g. thresholds, delays, stiffness) are rigid (constant), then the system may not be able to withstand any large inputs – i.e. may not adapt to the new circumstances. The only chance a system has to withstand strong environmental impacts, within bounds, is by varying some of its parameters as required by the occasion – e.g. to raise some of its thresholds, to lengthen some of its delays, or to decrease its stiffness, or to vary whatever other parameters in such a manner that its wholeness or integrity be preserved. A "plastic" system of this kind is capable of evaluating its own performance and adjusting some of its properties in such a way that its outputs be kept within the stability limits. Far from being a passive toy of its environment, such a system adapts itself actively to it and thus manages to endure. Because such systems are of paramount importance in the life sciences and in the social sciences, not to mention technology, they deserve a name of their own:

DEFINITION A1 A control system of the negative feedback type that changes some of its parameters in such a way that it remains stable in the presence of a changing environment is called a *self-governed, self-controlled, autonomic, adaptive*, or *plastic* system.

What happens if a system becomes unstable? There are two possibilities: if rigid it suffers *structural breakdown*, i.e. destruction. Nuclear disintegration and fission, the blowing up of a section of a network of any kind, the death of an organism, and the extinction of a biopopulation or an ecosystem, illustrate structural breakdown. On the other hand if the system is plastic it is said to suffer *functional breakdown*, i.e. it becomes a stable system of another kind. For example, an insulator subjected to a voltage above a certain point becomes a conductor; a man in a state of shock may lose some haunting memories; a society may become viable by modernizing its technology or getting rid of its parasites. In any case instability, usually dreaded by the engineer, is of interest to the scientist and the philosopher because it is often at the root of radical novelty. In other words, functional breakdown may result in structural buildup or morphogenesis.

2. GREY BOX MODELS

2.1. *Generalities*

Take a black box and observe or conjecture its internal states without however disclosing the mechanism that makes it tick: you have a grey box. Psychologists say that an *intervening variable* has been added to the stimulus-response pairs, and that the new variable is not a "hypothetical construct" but a mere formal link between inputs and outputs. But if the intervening variable is interpreted as representing the mechanism that transforms inputs into outputs, then instead of a grey box we have a dynamical box (Appendix B, Sec. 1). A grey box is then midway in depth between a black box and a translucid box or mechanismic model. We may picture it as a box equipped with a ladder every rung of which represents a state of the system: see Figure A7.

Of all the grey box theories the simplest is perhaps automata theory (Arbib, 1969; Ginsburg, 1962; Guinzburg, 1968; Harrison, 1965; McNaughton, 1961; Rabin and Scott, 1959; Rabin, 1963). As idealized by this theory, an automaton is a one component system admitting inputs of some kind (e.g. punched cards) and producing outputs of the same or of a different kind, such as printed symbols. The automaton jumps from one state to another in response to, or at least under the influence of, the inputs it accepts. All three – inputs, internal states, and outputs – are assumed to be denumerable. In short, an automaton is a discrete and sequential system. If lacking in spontaneity, i.e. acting only under external compulsion, the automaton is called *deterministic.* If inputs and internal states determine only output probabilities, the automaton is called *probabilistic.* Deterministic automata are of interest mainly in technology, where highly reliable systems are desirable. Probabilistic automata are of greater scientific and philosophical interest, for natural systems, such as brains and communities, are endowed with some spontaneity and seem to have strong stochastic components.

Automata are of greater interest than black box models because, unlike the latter, they take into account not only the behavior but also the internal states of the system. Consequently an automaton model can succeed (i.e. be true) where a black box model must fail (i.e. be false), namely (*a*) when a stimulus elicits different responses depending on the state of the system, or (*b*) when the system has spontaneous (uncaused) outputs. But of course automata models share with black box models the lack of concern with the specific nature of the thing modeled and even the spatial configuration of its components. In other words automata theory is a global (but not holistic), stuff-free (but not ghostly), atopic (nonspatial) and achronic (nontemporal) theory. Being so extremely poor, it is applicable to all

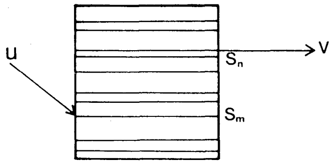


Fig. A7. Transition of system from state s_m to state s_n under the influence of input u.

kinds of system. (Perhaps for this reason von Neumann, 1951, mistook it for a chapter of formal logic.)

All information processing systems may be regarded, at least to a first approximation, as automata: nervous systems, servomechanisms, computers, TV networks, and so on. Therefore the referent of automata theory is not a narrow species of system but the entire genus of information processing systems, whether physical, chemical, living or social. And because the theory is concerned with certain traits of concrete systems, it is not a formal theory but a factual one. Indeed it provides an exact and simple (hence also superficial) model of a system interacting with its environment regardless of any specific features of interest to the special sciences – such as the kinds of material it is built from and the way it is energized. For this reason automata theory belongs not only in advanced technology but also in ontology. Let us therefore take a quick look at it.

2.2. *Deterministic Automata*

Automata theory presupposes only ordinary logic, the concepts of set and function, and some abstract algebra: it does not rest on any body of factual knowledge; in particular it contains no law statement borrowed from other theories. The specific defining (primitive) concepts of the theory of deterministic automata are collected in Table A1.

The above concepts are interrelated and made precise by the axiomatic

TABLE A1

Basic concepts of the theory of deterministic automata

Symbol	Mathematical nature	Factual interpretation
Σ	Set	Set of distinct unit inputs
Ω	Set	Set of distinct unit outputs
σ_0	Individual	Null input or pause
$\circ$	Binary operation	Concatenation
S	Set	State space
M	Function	Transition (next state) function
N	Function	Output function
s_0	Individual	Initial state
F	Set	Set of final states

DEFINITION A2 The structure $\mathscr{A} = \langle \Sigma, \Omega, \sigma_0, \circ, S, M, N, s_0, F \rangle$ represents a *finite deterministic automaton* (or *sequential machine with output*) A iff

(i) $\Sigma = \{\sigma_0, \sigma_1, \ldots, \sigma_k\}$, called the *input alphabet*, is a nonempty set with $k + 1$ elements called *letters*, every one of which represents a unit input on A from its environment;

(ii) Ω, called the *output alphabet*, is a set with two elements: 0 and 1, where 0 represents the absence of an output and 1 a unitary output;

(iii) let Σ^* be the set of finite concatenations, called *tapes*, of elements of Σ (i.e. letters), and let $\sigma_0 \in \Sigma$ be such that, for every tape $x \in \Sigma^*$, $\sigma_0 x = x\sigma_0 = x$. Likewise let Ω^* be the set of finite concatenations of elements of Ω, with $0 \in \Omega$ such that, for every $y \in \Omega^*$, $0y = y0 = y$. The structures $\langle \Sigma^*, \circ, \sigma_0 \rangle$ and $\langle \Omega^*, \circ, 0 \rangle$ are free monoids; $\circ$ represents the combination or concatenation of successive inputs to or outputs of A;

(iv) S, called the *state space* of A, is a finite nonempty set with n elements, where $n > 1$; every member of S represents an internal state of A;

(v) the function M: $S \times \Sigma^* \to S$ is called the *transition* (or *next state*) function; if $s \in S$ and $x \in \Sigma^*$, then $M(s, x) \in S$ represents the state A goes into when A, while in state s, admits input x;

(vi) the function N: $S \times \Sigma^* \to \Omega^*$ is called the *output function* and is such that $N(f, x) = 1 \in \Omega^*$ for every $f \in F \subset S$ and every $x \in \Sigma^*$, where $N(s, x)$ represents the output of A when acted on by the input $x \in \Sigma^*$ while in state s;

(vii) s_0 is a distinguished element of S representing the initial internal state of A;

(viii) F is a nonempty set included in the state space S, and every member f of F represents a final state of A;

(ix) A makes no spontaneous transitions; i.e. the null input has no effect: for every s in S, $M(s, \sigma_0) = s$;

(x) the internal states form sequences, i.e. the effect of a compound input $x \circ y$, where $x, y \in \Sigma^*$, equals the effect of the second input acting on the automaton in the state to which the first input carried it:

$$M(s, x \circ y) = M(M(s, x), y).$$

The last two assumptions are the essential components of Definition A2. They characterize an entire class of transition or next state functions, and state that a deterministic automaton operates only under external stimulation, sequentially, and in a determinate fashion. Nothing is said about the precise structure of the system, let alone its physics and chemistry: this is why automata theory has such a wide range of applications.

Examples The tables and state transition diagrams in Figure A 8 characterize two different automata with the same internal states reacting differently to the same unit inputs σ_1 and σ_2. They may thus represent some trait of the behavior of two different insects confronted with the same circumstances.

	s_0	s_1	s_2
$M_1(s, \sigma_1)$	s_1	s_2	s_1
$M_1(s, \sigma_2)$	s_2	s_0	s_2
$M_2(s, \sigma_1)$	s_0	s_2	s_2
$M_2(s, \sigma_2)$	s_1	s_0	s_0

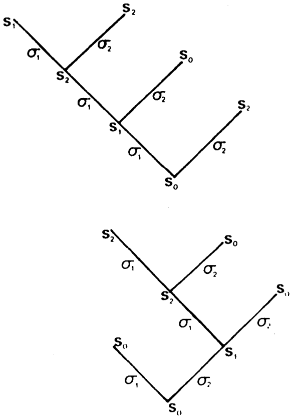


Fig. A8. Two different three state automata, each characterized by a different transition function (M_1 and M_2 respectively), subjected to the same stimuli.

In order to exhibit some consequences of the axiomatic definition of a deterministic automaton we need a couple of conventions. First,

DEFINITION A3 Let $\mathscr{A} = \langle \Sigma, \Omega, \sigma_0, \circ, S, M, N, s_0, F \rangle$ represent an automaton A. Then

(i) the restriction of the transition function M to s_0 is called the *response function*: $rp(x) = M(s_0, x)$, for $x \in \Sigma^*$;

(ii) a state $s \in S$ of A is *accessible* iff there is some input $x \in \Sigma^*$ such that $s = rp(x)$ – otherwise s is *inaccessible*;

(iii) a tape $x \in \Sigma^*$ is *accepted* (or *recognized*) by A iff $rp(x) \in F$ (i.e. if x has an overt effect when acting on A in s_0);

(iv) the *behavior* of A is the set of all the tapes accepted (or recognized) by A: $\beta(A) =_{df} \{x \in \Sigma^* \mid rp(x) \in F\}$;

(v) two automata A and A' are *behaviorally identical* iff $\beta(A) = \beta(A')$.

Note the difference between the above notion of behavior and that used in the so-called behavioral sciences.

And now a few typical consequences. Firstly

THEOREM A1 If two inputs $x, y \in \Sigma^*$ have the same effect, then the effects of the compound inputs xz and yz are the same for an arbitrary tape $z \in \Sigma^*$. I.e. for all states $s \in S$ and all inputs $x, y, z \in \Sigma^*$,

$$M(s, x) = M(s, y) \Rightarrow M(s, x \circ z) = M(s, y \circ z).$$

COROLLARY A1 For any tapes $x, y, z \in \Sigma^*$,

$$rp(x) = rp(y) \Rightarrow rp(xz) = rp(yz).$$

THEOREM A2 Every accessible state of an automaton which is the response to an input whose length is at least equal to the number of states of the system, can also be produced by a shorter input. In other words: If A has n states and if $s = rp(x)$ is an accessible state of A for some tape x composed of $m \geq n$ unit letters, then there exists another tape y such that $rp(y) = s$ and whose length is smaller than m.

THEOREM A3 Every accessible state of an automaton with n states is the response to some input of length smaller than n.

Let us now define the concept of automata analogy:

DEFINITION A4 Let $\mathscr{A} = \langle \Sigma, \Omega, \sigma_0, \circ, S, M, N, s_0, F \rangle$ and $A' = \langle \Sigma', \Omega', \sigma_0', \circ, S', M', N', s_0', F' \rangle$ represent finite automata A and A' respectively. Then

(i) A is *homomorphic* to A' iff there exists a mapping φ from $\mathscr{A}$ to $\mathscr{A}'$ such that

(*a*) $\varphi: \Sigma^* \cup S \to \Sigma'^* \cup S'$;

(b) for every u in $\Sigma^* \cup S$: if $u \in S$ then $\varphi(u) \in S'$;

(c) $\varphi(s_0) = s_0'$;

(d) for every t in $\Sigma^* \cup S$: if $t \in F$ then $\varphi(t) \in F'$;

(e) for every s in S and every σ in Σ: $\varphi(M(s, \sigma)) = M'(\varphi(s), \varphi(\sigma))$, $\varphi(N(s, \sigma)) = N'(\varphi(s), \varphi(\sigma))$;

(ii) A and A' are *isomorphic* iff φ is bijective;

(iii) A and A' are *structurally equivalent* iff they isomorphic;

(iv) the set of all automata isomorphic to a given automaton A is called the *structure type* of A.

It can be proved that, if there exists a homomorphism from one automaton to another, then it is unique. (The same holds, *a fortiori*, for isomorphisms.) The necessary and sufficient conditions for homomorphism and isomorphism are formulated in terms of

DEFINITION A5 An *equiresponse* relation $\sim_A$ on the set Σ^* of tapes is a binary relation on Σ^* such that, for any $x, y \in \Sigma^*$,

$$x \sim_A y \text{ iff } rp(x) = rp(y).$$

By using Definition A4 of homomorphism it is possible to prove

THEOREM A4 Let A and A' be automata. Then there is a homomorphism from A into A' iff the extension of $\sim_A$ is included in that of $\sim_{A'}$ and the behavior of A is included in that of A'.

COROLLARY A2 Two automata are isomorphic iff the extensions of their equiresponse relations coincide and their behaviors are the same.

2.3. *Probabilistic Automata*

The theory of probabilistic automata has the same formal background as that of deterministic automata plus elementary (discrete) probability theory. Its foundations are compressed into

DEFINTION A6 The structure $\mathscr{P} = \langle \Sigma, \Omega, \sigma_0, \circ, S, M, N, s_0, F \rangle$ represents a *finite probabilistic automaton P* iff all the coordinates of $\mathscr{P}$ but M satisfy the clauses of Definition $A2$, and M is now a function from $S \times \Sigma$ into the real interval $[0, 1]^{n+1}$, called the *table of transition probabilities*, such that, for every state $s \in S$ and letter $\sigma \in \Sigma$,

$$M(s, \sigma) = \langle p_0(s, \sigma), p_1(s, \sigma), \dots, p_n(s, \sigma) \rangle$$

where

(i) the $p_i(s, \sigma)$, for $i = 0, 1, \dots, n$, are probabilities, subject to the condition

$$\Sigma_i \, p_i(s, \sigma) = 1 \quad \text{for all } \sigma \in \Sigma;$$

(ii) $p_i(s, \sigma)$, for $i = 0, 1, \dots, n$, measures the strength of the disposition or propensity for the automaton P, when in state s, acted on by a unit stimulus σ, to jump into the state $s_i \in S$.

Since the transition function is now a probability, we have relinquished not only the determinate behavior of the system but also the serial or sequential character of its operations. The theory describes then the possible rather than the actual behavior of the system. How this is done can be gathered from what follows.

So far we have dealt only with unit inputs or letters. In order to deal with tapes or words we need one more item of notation:

DEFINITION A7 The *stochastic matrix* associated with a unit input $\sigma \in \Sigma$ is

$$P(\sigma) = \|p_j(s_i, \sigma)\|, \quad \text{with } i, j = 0, 1, \ldots, n.$$

The main result concerning the probability associated with a tape or word expresses the statistical independence of the effects of its components:

THEOREM A5 The stochastic matrix of an input $x = \sigma_p \sigma_q \cdots \sigma_r$ equals the product of the stochastic matrices of the unit components of the tape:

$$P(\sigma_p \sigma_q \cdots \sigma_r) \equiv \|p_j(s_i, \sigma_p \sigma_q \ldots \sigma_r)\| = P(\sigma_p) \cdot P(\sigma_q) \cdots P(\sigma_r).$$

DEFINITION A8 Let $\mathscr{P} = \langle \Sigma, \Omega, \sigma_0, \circ, S, M, N, s_0, F \rangle$ represent an automaton P and let $I = \{i_0, i_1, \ldots, i_r\} \subset \mathbb{N}$ be an index set used to label the final states, i.e. set $F = \{s_{i_0}, s_{i_0}, \ldots s_{i_r}\}$. The *probability of tape* $x \in \Sigma^*$ *being accepted* (recognized) by P when in s_0 is

$$p(x) = \sum_{i \in I} p_i(s_0, x).$$

Example Let $\Sigma = \{\sigma_0, \sigma_1, \sigma_2\}$, $S = \{s_0, s_1\}$, $F = \{s_1\}$, and assume the stochastic matrices

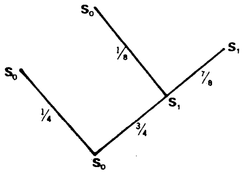
$$P(\sigma_1) = \left\| \begin{matrix} 1 & 0 \\ \frac{1}{2} & \frac{1}{2} \end{matrix} \right\| \qquad P(\sigma_2) = \left\| \begin{matrix} \frac{1}{2} & \frac{1}{2} \\ 0 & 1 \end{matrix} \right\|.$$

Then the table of transition probabilities for the tape $x = \sigma_1 \sigma_2 \sigma_2$ is

$$P(\sigma_1 \sigma_2 \sigma_2) = P(\sigma_1) \cdot P(\sigma_2) \cdot P(\sigma_2) = \left\| \begin{matrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{8} & \frac{7}{8} \end{matrix} \right\|$$

$$= \left\| \begin{matrix} p_0(s_0, x) & p_1(s_0, x) \\ p_0(s_1, x) & p_1(s_1, x) \end{matrix} \right\|.$$

The corresponding diagram in the state space is



Only the probable set of behaviors of a probabilistic automaton can be computed: its actual behavior can only be described after the event. However, if the various transition probabilities are markedly different from one another, it is possible to anticipate the most probable behavior of an automaton by taking into account only those inputs that have a fair chance of being accepted, and ignoring all others. More precisely, one can choose a real number comprised between 0 and 1, called the *cutpoint*, below which all probabilities can be neglected – at a risk of course. In other words, one can use

DEFINITION A9 Let P be a probabilistic automaton and let r be a real number such that $0 \leq r \leq 1$. Then *the most probable behavior* above (the cutpoint) r is

$$\beta_r(P) = \{x \in \Sigma^* \mid p\,(x) > r\}.$$

The relation between probabilistic and deterministic automata, as regards their behavior, is given by

THEOREM A6 P is a deterministic automaton iff $\beta_r(p) = \beta(P)$ for any r such that $0 \leq r < 1$.

This is but an instance of an important metatheorem that has no known proof, namely: *Every stochastic theory includes a deterministic subtheory.*

2.4. *Information Systems*

Our last example of grey box models will be the information system. The signals impinging upon an automaton, whether deterministic or probabilistic, drive it from one state to the next, either invariably or with a certain probability, without altering the set of its possible final states. On the other hand if certain inputs can alter – e.g. expand or shrink – the subset of final states in the state space of a system, then the latter does not qualify as an automaton but does qualify as an *information system*. That is, an information system is one that can gain or lose final states in response to certain external stimuli. If the latter have such an effect, they are called *signals* or *messages*, and each such message can be assigned a definite amount of information measuring the effect of the message on the system.

Needless to say, every message is carried by some information carrier or channel, and every information gain or loss involves some energy. However, the general theory dealing with information systems is not interested in the physical characteristics of either system or message, not even in the energetic aspect of the transaction: it focuses instead on the net change in the number of final states (or outcomes) of a system affected by a message, whatever its nature. So much so that it regards source, channel and receiver as atomic parts of the system and does not include any physical variables. See Figure A9. (In the popular literature on information it is often stated that an input to a system is called a *signal* iff it elicits an output whose energy is much greater than the input energy. However, information theory does not contain the concept of energy, so the popular characterization is inadequate.)

Consider then a system with n possible final states and a message arriving at the system, the effect of which is to reduce that number of alternatives from n to some number $m \leq n$. The *amount of information* carried by the message is taken to equal $\log_2(n/m)$ bits. (The *bit* is of course the unit of information, i.e. the quantity of information gained by the system when the number of alternatives is halved. In fact, in this case $\log_2(n/(n/2)) = \log_2 2 = 1$.) In short, if a message shrinks the number of final states of a system from n to $m < n$, it is said to carry the quantity of information $I = \log_2 (n/m)$. There need be nothing cognitive, let alone subjective, about this concept of information.

If the signals are random, either because they are generated at random or, more commonly, because the information channel is subject to random disturbances, then the statistical theory of information (Shannon and Weaver, 1949) applies. The basic notion of this theory is that of a random binary physical signal appearing at the receiving end of a communication system. The central postulate of the theory is that the quantity of information of such a signal is $I = \log_2(1/p)$, where p is the probability of the signal's reaching the receiver. If the latter happens to be an intelligent being then I is interpreted as the degree of unexpectedness of the message. But

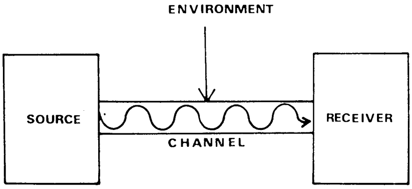


Fig. A9. The components of an information system. The effects of the environment on the channel, when random, are lumped under the category of noise.

this interpretation, legitimate as it is in some applications of statistical information theory to psychology, is irrelevant to communications engineering, molecular biology, and semantics.

One final note of caution. Some claim that information measures organization. The argument runs like this: a deterministic system is predictable, so watching its behavior provides no information; on the other hand a probabilistic system is unpredictable, so whatever it does is surprising or informative. Hence the more disorganized a system the more informative the observation of it. This argument contains several important mistakes. Firstly, the information function is not defined on a set of systems (or of states of a system) but on system-message pairs. Consequently a quantity of information, far from being an intrinsic property of a system (e.g. a transfer RNA molecule), is a mutual property of system and message. Hence it cannot measure the degree of organization of a system nor any other intrinsic property of it. Secondly, systems are neither predictable nor unpredictable: whatever can be predicted can be so with the help of theories and data. The predicate "predictable" is not defined on the set of systems but on the set of pairs ⟨body of knowledge, system⟩. Thirdly, prediction must not be equated with deterministic prediction. If a system is stochastic, as is the case with an atom, a learning organism, or a social group, then its behavior may be predicted statistically – provided we have succeeded in building a true probabilistic theory about it. In this case we may be able to predict probabilities (of e.g. transitions) and a number of properties derived from them, such as averages and standard deviations. Fourthly, a stochastic system – such as a probabilistic automaton – is not chaotic: being probabilistic it is "governed" by probabilistic laws. Thus, though stochastic, an atom is not chaotic like the top of my desk, over which no probability distribution can be defined. (See Vol. 3, Ch. 4, Sec.6.4 for the difference between randomness and chaos.)

In conclusion, information does not measure organization. And there seems to be no better measure, perhaps because organization is an extremely complex property that defies lumping into a single number. What we do have are measures of efficiency, such as thermodynamic efficiency, output/input ratios, or signal/noise ratios. Knowing that a system is highly (or fairly or poorly) efficient we can infer that it is well (or fairly or poorly) organized to attain a certain goal or to help us perform a certain task. But efficiency is an indicator of organization not a measure of it.

APPENDIX B

CHANGE MODELS

In this Appendix the accent will be on change. We shall exhibit both kinematical and dynamical models, first of quantitative, then of qualitative change. And we shall study only extremely general change models, i.e. such that are portable from one field of inquiry to another. As in the case of Appendix A, most of the material to be presented is standard or has at least appeared before. But the interpretation and the philosophical discussion is often novel.

1. KINEMATICAL MODELS

1.1. *Global kinematics*

A kinematical model of a system is a theory representing the change, step by step, in some of the properties of the system. The simplest kinematical model assumes that the properties of the system are representable by smooth functions, and that the collection of all of these, i.e. the state function, is a time dependent function with values in $\mathbb{R}^n$ obeying an equation of evolution of the form

$$\dot{\mathbb{F}} = \mathbb{G}(\mathbb{F}, t).$$

Example The rate of change of the state function is a linear combination of its own components, i.e. $\mathbb{F} = G\cdot\mathbb{F}$, where G is a square matrix with time independent coefficients. The solution is $\mathbb{F}(t) = e^{Gt}\cdot\mathbb{F}(0)$, where $\mathbb{F}(0)$ is the value of the state function at $t = 0$.

Some of the components of $\mathbb{G}$ may remain constant throughout the evolution of the system. In this case they are called the (time) *invariants*, or *constants of the motion*, of the system. (The system as a whole need not move through space.) If all the n components of $\mathbb{G}$ vanish, the system has n constants of motion and is always in a steady state even though its components may not be in such a state.

Another systemic or global property that can be studied with the help of a kinematical model is stability. A system represented by a state equation of the form $\dot{\mathbb{F}} = \mathbb{G}(\mathbb{F}, t)$ is (a) in a *stationary state* (or in *equilibrium*) during a certain time interval iff $\dot{\mathbb{F}} = 0$ during that interval (so that $\mathbb{F}$

equals a constant n-tuple); (*b*) *asymptotically stable* iff $\mathbb{F}$ approaches an equilibrium value for increasing values of t.

Example Suppose a system is characterized by just two properties, represented by F_1 and F_2, so that its equations of evolution are

$$\dot{F}_1 = G_1(F_1, F_2, t), \qquad \dot{F}_2 = G_2(F_1, F_2, t).$$

The equilibrium property can be found by going over to the state space, namely by dividing one equation by the other to obtain

$$\frac{\mathrm{d}F_1}{\mathrm{d}F_2} = \frac{G_1}{G_2},$$

which represents the possible processes in the state space. The system is stable iff $\mathrm{d}F_2/\mathrm{d}F_1$ approaches a definite number as F_1 increases. This number is the slope of a straight line in the state space, namely the equilibrium trajectory. If on the other hand the system diverges from the (imaginary) equilibrium line, it is unstable. See Figure Bl. (The equilibrium trajectory may be a finite segment or just a point.)

1.2. *Analytical Kinematics*

In the preceding subsection we dealt with the global kinematics of an arbitrary system. In contrast, analytical kinematics pays attention to the components of a system and the links among them. Assume, for the sake of simplicity, that a system has just two components characterized by one property each. Call F_1 and F_2 the functions representing the peculiar property of component 1 and component 2 respectively. Suppose that these functions vary smoothly in the course of time and that the law of evolution is

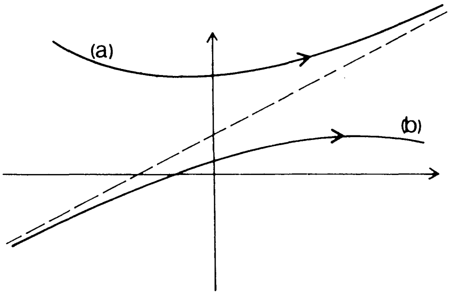


Fig. B1. (*a*) Stable system. (*b*) Unstable system.

$$\dot{\mathbb{F}} = \mathbb{G}(\mathbb{F}) \quad \text{with} \quad \mathbb{F} = \langle F_1, F_2 \rangle, \quad \mathbb{G} = \langle G_1, G_2 \rangle.$$

Assume further that the system is linear, i.e. that

$$\dot{\mathbb{F}} = A \cdot \mathbb{F} \quad \text{with} \quad A = \begin{Vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{Vmatrix}, \quad \text{where } a_{ii} \in \mathbb{R}.$$

If the off-diagonal coefficients vanish or are very small compared with the diagonal elements, then each component evolves separately or nearly so and, instead of a system proper, we have just an aggregate. Otherwise the evolutions of the system components are coupled.

Example Let $F_i(t)$ represent the amount of stuff of some kind in the ith component of the system at time t. (Such stuff can be matter, information, goods, capital, or what have you.) The system will be in a stable state, and remain in it, provided each component gives back as much stuff as it receives, i.e. iff $a_{12} = a_{21}$. Any departure from this equilibrium state will ensue in an uneven distribution of the stuff measured by F_i – i.e. there will be wind, a leak, an imbalance of payment, a migratory flow, or some other diffusion or transport process. If this trend continues unchecked – i.e. if the system lacks control mechanisms – the entire stuff may end up by concentrating in one of the components. At this point of maximal heterogeneity the exchange may cease and thus the system itself will come to an end *qua* system: each component starts evolving independently from the other. In short, depending on the control mechanisms, whatever disparity there was will either level out (as is the case with heat transfer) or be greatly enhanced (as is the case with unchecked capital transfer). In short, if balance is required then checks are called for.

1.3. *Balance Equations*

So far in this Appendix we have not singled out any state variable representing the transactions of the system with its environment. We shall now remedy this shortcoming. To this end we shall have to introduce the concept of ordinary (physical) space.

Example In $\dot{\mathbb{F}} = \mathbb{G}(x, t)$ set $\mathbb{G} = - \nabla \cdot j$, where j is a vector valued function of position and time representing the flux of something. This law schema occurs so frequently under different guises that it is called the *general balance* (or *conservation* or *continuity*) equation, and is sometimes regarded as a purely mathematical statement without any factual content. Let us examine this matter.

Let σ be a system with volume V and surface S in Euclidean three-space. (We are assuming that σ is not a megasystem, so it does not require Riemannian geometry.) Call ρ an intensive (nonextensive) property of ε, and assume that σ interacts with its environment. By definition, the extensive property ε associated with ρ is

$$\varepsilon(\rho, t) = \int_V dV \rho(\sigma, x, t).$$

For the sake of definiteness think of ρ as the mass density and of ε as the total mass of the system.

We now assume the following factual hypothesis:

POSTULATE B1 For every extensive property ε of a system interacting with its environment, ε may change in the course of time in two (mutually compatible) ways: spontaneously (internally) or as a consequence of system-environment interactions. More precisely, the rate of change of ε is composed of two parts: a source s and a flux j, or

$$\dot{\varepsilon} = \int_V \mathrm{d}V s + \int_S \mathrm{d}S\, j_n,$$

where j_n is the projection of j onto the interior normal to S. See Figure B2.

Using now Green's theorem, a purely mathematical result, we derive

THEOREM B1 If ρ is the intensive property associated with the extensive property ε occurring in Postulate B1, then

$$\dot{\rho} + \nabla \cdot j = s.$$

This is the local (rather than integral or global) balance equation. It is not a special law statement but a *law schema*: it holds for every intensive

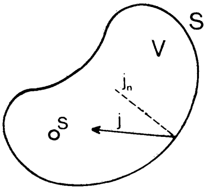


Fig. B2. Transactions of a system with its environment as reflected in the rate of change of the former's extensive property ε. This rate consists of the (positive or negative) contribution of the source (or sink s) and the (positive or negative) contribution of the flow j.

property of an arbitrary system, ergo it will occur as often as intensive properties are treated, whether in quantum mechanics or in continuous models of traffic flow. (The only limitation on it is that it is local, i.e. it holds only in the neighborhood of a reference frame.)

Example Let $j = -a(\partial\rho/\partial x) - b\rho$, with a and b real numbers, and $s = 0$. Then

$$\dot{\rho} = a\,\partial^2\rho/\partial x^2 + b\,\partial\rho/\partial x$$

represents the superposition of a diffusion process, that may be random, and a drift caused by the force $b\,\partial\rho/\partial x$. But even this is just a law schema unless ρ is interpreted in factual terms.

Theorem B1 allows us to give a quantitative definition of the notion of system closure:

DEFINITION B1 Let a system σ interacting with its environment possess n intensive properties ρ_i with the corresponding fluxes j_i, where $1 \leq i \leq n$. Then σ is *isolated* from (*closed* to) its environment during a given time interval

(i) *in the ith respect* iff the flux j_i vanishes on the system boundary for all time in the given interval;

(ii) *in all respects* iff all the fluxes j_i vanish on the system boundary for all time in the given interval.

COROLLARY B1 For any system isolated (closed) in the ith respect, $\dot{\rho}_i = s$.

Moral: Before blaming any changes in the system on its environment, search for the system's internal sources of change. This holds not only in the natural sciences but also in the social ones.

1.4. *Lagrangian Framework*

Consider a system of any kind with n components each of which is characterized by just three basic state functions (called *generalized coordinates*), the values of which depend only on a phase variable interpretable as the time coordinate of a certain reference frame. A Lagrangian model of such a system boils down to a certain function L of the $3n$ generalized coordinates, their rates, and time, that satisfies a certain extremality condition. Because the nature of the components is left unspecified, the model fits any system with a finite number of distinct parts, be it physical, biological, or social. This referential generality explains the wide use of lagrangian kinematics in a variety of scientific and technological fields.

A possible formalization of the notion of a lagrangian model is given by

DEFINITION B2 Let σ be a system composed of n parts and moreover one with a constant composition, i.e. such that $\mathscr{C}_t(\sigma) = \mathscr{C}_{t'}(\sigma)$ for any t, $t' \in T$. Then the structure $\mathscr{L} = \langle T, Q, L \rangle$ is a *lagrangian model* of σ iff

(i) T is a subset of the real line ordered by the relation $\leq$ of temporal priority or simultaneity, and every $t \in T$ represents an instant of time;

(ii) $Q = \{q_i \mid q_i\colon T \to \mathbb{R} \;\&\; 1 \leq i \leq 3n\}$ is a set of time dependent, real valued and once differentiable functions [the basic state functions of the system];

(iii) the $6n$-tuple $\langle q_1(\sigma, t), \ldots, q_{3n}(\sigma, t), \dot{q}_1(\sigma, t), \ldots, \dot{q}_{3n}(\sigma, t) \rangle$ represents the state of σ at time t;

(iv) L is a real valued function of the state variables, their rates, and t – i.e. $\langle q_i(t), \dot{q}_i(t), t \rangle \mapsto L(q_i, \dot{q}_i, t)$ – such that, for any $t \in [t_1, t_2] \subset T$ and fixed end points $q_i(t_1)$ and $q\,(t_2)$, the integral

$$I =_{df} \int_{t_1}^{t_2} \mathrm{d}t\, L\,(\sigma, q_i, \dot{q}_i, t)$$

is either minimal or maximal.

By performing the indicated first order variations (i.e. by computing δI and setting it equal to 0) it is easily seen that the above axiomatic definition entails the following system of $3n$ equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad 1 \leq i \leq 3n.$$

If there are constraints, i.e. if the generalized coordinates q are not mutually independent, the equations of "motion" are more general: the right hand side equals a function Q_i representing the net effect of the constraints.

Sometimes these kinematical equations can be assigned a dynamical interpretation, namely as representing a balance of forces. In particular $\partial L/\partial q_i$ represents the intensity of the combined action of the environment and the rest of the system upon the component characterized by three values of the index i, and Q_i the forces amounting to the constraints. The latter are particularly important in the case of social systems, where no variable can be said to be independent of the others.

Note that so far we have not specified the Lagrangian function L. A suitable choice of L and of the semantic assumptions concerning the generalized coordinates q_i will yield the laws of particle mechanics. An alternative interpretation yields the theory of electrical networks. Still another interpretation ensues in a certain theory of population dynamics – and so on. In sum, the lagrangian formalism is subject-matter independent. For this reason it belongs not only to science and technology but also to ontology (Bunge, 1974c).

In many fields of inquiry the equations of evolution are of the form

$$\dot{\mathbb{F}} = \mathbb{G}(\mathbb{F}, t)$$

(Sec. 1.1). These equations are just a particular case of the lagrangian formalism with constraints. In fact regarding the components F_i of the state function $\mathbb{F}$ as the generalized coordinates, and introducing the definitions

$$L =_{df} \sum_{k=1}^{n} F_k \dot{F}^k - \sum_{k=1}^{n} G_k F^k$$

$$Q_i =_{df} \mathbb{F} \cdot \frac{\partial \mathbb{G}}{\partial F_i},$$

the two sets of equations coincide. (And, because $\partial L/\partial \dot{F}_i = F_i$, the generalized momenta $p_i = \partial L/\partial \dot{F}_i$ are seen to coincide with the corresponding generalized coordinates. This does no harm since, in general, the generalized coordinates are not interpretable as position coordinates. Thus in the case of a society F_1 may represent the population, F_2 the energy consumption per capita, F_3 the GNP per capita, and so on.) So much for the Lagrangian models of discrete systems.

In the case of a continuous system, such as an elastic body, an electric field, or a sufficiently dense population of organisms, one starts from a lagrangian density and obtains L as the volume integral of that density. As with the discrete case, a few rather mild conditions on the Lagrangian suffice to yield a theory schema with a wide coverage – so wide indeed that, when enriched with suitable subsidiary conditions or hypotheses, it yields every one of the classical field theories (Edelen, 1962). Such conditions specify then a specific Lagrangian model.

What is usually called *axiomatic field theory* intends to discharge a similar function with respect to the quantum field theories, in particular quantum electrodynamics. However, unlike the general classical field theory mentioned a moment ago, its quantum counterpart does not yield quantum electrodynamics as a special case. "It can at most give a framework for a concrete theory of existing particles and their corresponding fields" (Jost, 1965, p. xi). In addition, one hopes that this approach may avoid the inconsistencies that mar the quantum field theories. Its role in metaphysics is yet to be investigated.

We close this subsection with an historical remark. The Lagrangian framework was foreshadowed by Maupertuis as a synthesis of dynamics and metaphysics. It was brought to fruition by Lagrange, who freed it from metaphysics, and it was generalized beyond dynamics by Maxwell. The axiom contained in Definition B2 (iv) is one of the most secure and yet misinterpreted principles of science. Maupertuis took the (action) integral I to be always a minimum and interpreted this as a sure sign of divine wisdom: Nature always chose the paths that minimized I. It was later discovered that in some cases (notably in optics) I is not a minimum but a maximum, so the principle was rechristened from that of minimal to that of extremal action. Nevertheless a number of physicists, among them Planck and Margenau, continued to regard Maupertuis' principle as an expression of purpose and design. This is mistaken if only because the principle, together with some lemmas borrowed from the variational calculus, entails the equations of motion – in which no traces of purpose or design have ever been detected. It is not that the system behaves in order to minimize or maximize its action: it is just a law, and an extremely pervasive one, that all systems behave that way. Because of this generality the Lagrangian framework should be recognized as belonging to the intersection of science and metaphysics.

1.5. *Kinematical Analogy*

We close this section with some remarks on the notion of kinematical analogy. Any systems which, regardless of their detailed composition and structure, have the same kinematics – hence are described adequately by the same equations of evolution – can be said to be *kinematically analogous* (or *equivalent*). For example some periodic oscillations, whether mechanical, electrical, chemical, biological, or economic, happen to satisfy the same differential equations, hence they are kinematically analogous. Thus the relaxation oscillations in a vacuum tube amplifier are described by the equation $\ddot{x} - a(1 - x^2)\dot{x} + bx = 0$, which has been said to describe also the noise made by scratching a piece of chalk on a blackboard, the waving of a flag in the wind, shivering from cold, and the periodic recurrence of epidemics.

There can be no doubt that the finding of kinematical analogies can play a heuristic role and save time. But they have hardly any ontological import except insofar as they show that different mechanisms may result in similar overall behavior patterns, whence the study of behavior does not exhaust that of reality. For example brownian particles, domestic flies, and drunkards move randomly, but the underlying mechanisms are very different, so that the observation that all three are kinematically analogous has no metaphysical significance. What does have ontological import is the finding of *dynamical* analogies among systems of different kinds, i.e. the discovery that certain *mechanisms* – such as transport, diffusion, merging, cooperation, competition, random shuffling, and casual mutation – are conspicuous on all levels. So, let us turn our attention to mechanisms.

2. DYNAMICAL MODELS

2.1. *Generalities*

A dynamical model of a system is a kinematical model (Sec. 1) some variables of which are interpretable as "driving" others. In other words, a dynamical model involves a change *mechanism*, i.e. either a subsytem or a process whereby the changes in some properties of the system effect changes in some other properties of the same system. Morever all such changes are spatiotemporal or reducible to changes in time (such as rates) or in space (such as gradients).

Example 1 The geographic distribution of a biospecies can be explained as the outcome of the action of three separate mechanisms: random diffusion, drift, and environmental resistance (or selection pressure). A dynamical model of the spatial pattern of a species (or rather a family of populations of the given species) is this. The rate of change of the population density ρ of the given collection of populations at some point x on the surface of the planet is

$$\frac{\partial \rho}{\partial t} = \underset{\text{Diffusion}}{\alpha\, \partial^2\rho/\partial x^2} - \underset{\text{Drift}}{2\beta\, \partial\rho/\partial x} + \underset{\substack{\text{Environmental}\\ \text{pressure}}}{\gamma\rho(\rho - 1)}.$$

Example 2 Much the same holds if ρ is interpreted, not as the actual population density, but as the probability density, so that $\rho\, \mathrm{d}S$ represents the probability that $\rho\, \mathrm{d}S$ individuals live in the area $\mathrm{d}S$. In this case all three forces are stochastic, like the pushes exerted by molecules on a brownian particle. *Example 3* Field models are a kind of dynamical model. A field is a thing spread in a continuous manner over an expanse of space. (Think of the gravitational or of the magnetic field of our planet.) In principle a field does not qualify as a system any more than a particle does, for neither has atomic parts. However, a collection of fields superposed in a given region of space does qualify as a system. An electromagnetic wave is an example of a pure field system. Also, a field may play a part in a system not made up exclusively of fields: its role is that of glue that keeps the nonfield components together. We call a *field theoretic model* of a system any model in which field concepts occur essentially, as is the case with the (theoretical) model of a molecule. The "fields" invented by some biologists, psychologists, and social scientists, are phoney, for they are not substantial continua and are not described by field equations: they are just metaphors.

A field theoretic model is mechanismic, not phenomenological, but its mechanism is nonmechanical. (Mechanical mechanisms are a proper subset of mechanisms.)

2.2. *Formalities*

We repeat in a slightly more exact way the definition of a mechanism given at the beginning of the last subsection:

DEFINITION B3 Let σ be a system. Then a *mechanism* of σ is either
(i) a component m of σ such that m acts in σ, or
(ii) a process m in σ such that other processes in σ depend upon m.

And now we assume

POSTULATE B2 Every system has some mechanism.

Clearly this postulate (*a*) overreaches experience, (*b*) goes against the grain of phenomenalism and descriptivism (e.g. behaviorism), and (*c*) encourages theorizing. The kind of theorizing our postulate stimulates is, of course, the building of dynamical models. A *dynamical model* of a system may be defined as a model representing one or more mechanisms of the system.

A standard formulation of a dynamical model is within the general framework of hamiltonian dynamics. Just as every kinematical model can be formulated in Lagrangian terms (Sec. 1.4), so nearly every dynamical model can be formulated in a Hamiltonian fashion. It will be recalled that the former has just one set of $3n$ basic state functions, namely the generalized coordinates q_i. A Hamiltonian model, on the other hand, rests on two sets of basic state functions: the very same generalized coordinates q_i and the corresponding generalized momenta p_i, which are logically independent of the former. Just as the solutions q_i of a Lagrangian problem may represent the spatial configuration of the system, so the corresponding p_i represent the dynamics of the system. In particular, the rates $\dot{p}_i$ may represent the forces acting on the various parts of the system. If the Lagrangian L of the system is known, then the p_i are given by $p_i = \partial L/\partial\dot{q}_i$, and the hamiltonian of the system by $H = \Sigma p_i\dot{q}_i - L$. But if L is unknown then the problem must be formulated afresh by conveniently choosing the basic variables and guessing the adequate Hamiltonian. The basic equations of motion are

$$\dot{p}_i = -\partial H/\partial q_i, \qquad \dot{q}_i = \partial H/\partial p_i.$$

So, the solution of a hamiltonian problem is compressed into the state function $\mathbb{F} = \langle q_1, q_2, \ldots, q_{3n}, p_1, p_2, \ldots, p_{3n}\rangle$. Every value of this function represents a possible state of the system. And as time goes by the tip of $\mathbb{F}$ traces a trajectory in the state space of the system, which in this case is called its *phase space.*

Although the general hamiltonian theory – namely that in which H is left unspecified and the basic state variables uninterpreted – belongs to scientific ontology, we shall not go into any details of it. (For this and a few other general dynamical theories, see Bunge, 1967b, Ch. 2.) Our purposes in mentioning the theory here are (*a*) to emphasize the richness of the ontology that originates in science, and (*b*) to elucidate the concept of dynamical analogy. The latter may be defined as follows. Let σ and σ' be two systems modeled by the Hamiltonians $H_\sigma(q, p, t)$ and $H_{\sigma'}(q', p', t')$ respectively. Then σ and σ' are said to be *dynamically* analogous iff H_σ and $H_{\sigma'}$ are formally the same up to a similarity transformation of its dependent variables – i.e. if $q' = aq$, $p' = bp$, and $t' = ct$, where a, b and c are complex numbers. Nothing is required of the interpretation of these variables.

Let us next discuss the dynamics of processes of two special yet rather widespread kinds, namely competition and cooperation, which will be treated jointly in a single dynamical model.

2.3. *The Pervasiveness of Cooperation and Competition*

When two or more things get together and form a system, they may be said to cooperate with one another even if they do not do so deliberately.

For example, each of the syntheses (physical, chemical, biological, or social) $A + B \rightharpoonup C$ and $A + D \rightharpoonup E$ may be regarded as a cooperative process: the A's cooperate with the B's to form C's, and likewise for the second reaction. However, if the two processes occur simultaneously at the same place, then the two cooperative processes compete with each other: indeed, in this case the B's and the D's compete for the A's. So, in this case there is both cooperation (in some respects) and competition (in others). Likewise two families (or other sociosystems) may compete with one another while cooperating internally.

A possible definition of the notions of interest is this:

DEFINITION B4 Let x, y and z be things. Then

(i) x *cooperates* with y iff there is a system z such that x and y are in the composition of z;

(ii) x *competes* with y for z iff the formation of a system composed of x and z interferes with that of a system composed of y and z.

Note that, according to this definition, competition is not the opposite of cooperation. Moreover two things may cooperate in certain respects and compete in others.

Cooperation has received far less attention than competition. In particular, most philosophers since Heraclitus have stressed war at the expense of cooperation, ignoring that intersystemic competition calls for intrasystemic cooperation. On the other hand biologists, pharmacologists and physicians are fond of emphasizing that the components of the normal cell act in cooperation, that every healthy organism is synergic, and that only sickness and senility cut the cooperative ties.

Of late physicists have joined life scientists in studying a number of cooperative processes. One striking discovery along this line marked the birth of a whole new branch of science, namely plasma physics. Collective or molar modes of oscillation and propagation were discovered in such lowly systems as bodies constituted by very hot and highly ionized gases – which, according to classical ideas, should be the paradigm of chaos and homogeneity. Plasma waves, both stationary and traveling, as well as the emergence of complex shapes of plasma bodies (plasmons), gave the lie to the belief that in every gas all a molecule can do is to move haphazardly and collide with other molecules in a random fashion without ever giving rise to a new system characterized by novel overall modes of behavior. At about the same time chemists discovered the hydrogen bond, which can give rise to marvels of cooperation such as the biomolecules. In short, cooperation is nowadays just as "in" as competition was one century ago. Which is just as well, for unless cooperation is acknowledged alongside competition, the very existence of systems cannot be accounted for. For one thing the pervasiveness of assembly processes would not have been recognized if scientists had adhered to the Heraclitean ontology of strife.

We shall stick our neck out and assume

POSTULATE B3 Cooperation is about as frequent as competition.

We shall presently atone for the vagueness of this formulation by reviewing a mathematical model of processes wherein cooperation intertwines with competition (Bunge, 1976).

2.4. *The Dynamics of Competitive–Cooperative Processes*

Suppose that two things or systems, *a* and *b*, consume or produce a third thing *c*. We make no assumption concerning the nature of these things, let alone the mechanism of consumption or production. But we do assume something about the modes of cooperation and competition between *a* and *b* with regard to *c*. In fact we assume that each component, when on its own, consumes or produces *c* at a constant rate (possibly nil), and that each may or may not succeed in making ends meet. We also assume that the components are sure to attain the optimum level (of consumption or of production) if they cooperate with one another. Finally, we suppose that this cooperation becomes intenser the more it is needed, that it ceases when no longer needed, and that from then onwards it turns into hindrance, so that growth remains under control. Such a process can be formalized as follows.

Let $A(t)$ and $B(t)$ be the instantaneous amounts of thing *c* consumed or produced by the components *a* and *b* respectively, and call α and β the optimum or satiety levels. Further, let

$$A(0) = c_1 < \alpha, \qquad B(0) = c_2 < \beta$$

be the initial values. In the absence of interaction the rates are assumed to be constant (possibly zero):

$$\dot{A} = a_{11}, \qquad \dot{B} = a_{22} \quad \text{with} \quad a_{11}, a_{22} \in \mathbb{R}^{+}.$$

If these rates are nil, no processes occur. If they are positive, each individual eventually attains and surpasses its optimum level. Furthermore if cooperation sets in from the start, *a* will help *b* in proportion to the latter's needs, i.e. in such a way that the imbalance $\beta - B$ will be offset; and b in turn contributes in proportion to the imbalance $\alpha - A$, though not necessarily with the same intensity. This mutual help, we have assumed, is positive as long as there is a deficit (negative balance); it becomes nil on attaining satiety, and negative as soon as there is superavit (positive balance). The simplest formulas representing such a cooperative process are the nonlinear equations

$$\dot{A} = a_{11} + a_{12}(\alpha - A)B, \qquad \dot{B} = a_{22} + a_{21}(\beta - B)A$$

with $a_{ij} \geq 0$ for $i, j = 1, 2$.

The coefficient a_{12} measures the strength of the help a gets from b, while a_{21} measures the intensity of the help a lends b. At the optimum point $\langle \alpha, \beta \rangle$ there is neither help nor hindrance: each component fends for itself, so the system breaks down momentarily.

We spare the reader the treatment of the general case in the vicinity of the optimum point, showing that, in effect, no sooner does a attain the α level than its partner b turns against it forcing it to slow down its rate of consumption or of production. (Cf. Bunge, 1976.) Instead, we shall study the particular case when the two partners behave in exactly the same way. In this case, i.e. for $c_1 = c_2 = c$ (no initial advantage), $a_{11} = a_{22}$ (same individual rates), $a_{12} = a_{21}$ (symmetric help) and $\alpha = \beta$ (same requirements), we are left with a single equation of evolution, namely

$$\dot{X} = a_{11} + a_{12}(\alpha - X)X, \qquad X = A, B.$$

The general solution to this equation is found to be

$$X(t) = a_{11}t + \frac{\alpha c}{c + (\alpha - c)\exp(-a_{12}\alpha t)}.$$

If each partner is self-sufficient ($a_{11} = a_{22} > 0$), cooperation only speeds up the process up to the optimum point, and curbs it after that point. Otherwise (i.e. if $a_{11} = a_{22} = 0$), mutual help makes the process possible. Indeed, the second term of the last equation represents the growth of A (which is equal to B) from the initial value c towards the asymptote $A(\infty) = \alpha > c$. Rivalry from this point on prevents unlimited growth in either consumption or production. See Figure B3. This limitation takes place whether or not there is an external limitation such as scarcity of resources: the control is strictly internal. Not that there are no external limitations; only, our model does not take them into account.

General dynamical models of this kind have at least three possible uses:

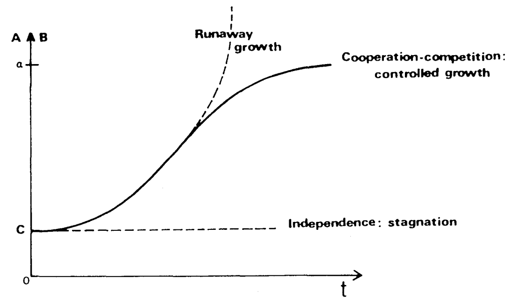


Fig. B3. Full line of a system composed of partners that cooperate and compete ($a_{12} \neq 0$) but are powerless in isolation (i.e. when $a_{11} = a_{22} = 0$). Dotted line: stagnation in the absence of interaction (i.e. when $a_{12} = 0$).

(*a*) they may suggest the search for real systems (e.g. in morphogenesis, neurology, ecology, or sociology) that approach a zero growth state in some variable as a result of their cooperative-competitive structure; (*b*) they may stimulate the building of specific theories covering real systems behaving in the way outlined by the model; (*c*) they may help in planning social controls. In particular, our model suggests redesigning society in such a way that each member helps his/her neighbor in proportion to his/her own ability and the other party's needs, and also controls his/her partner's growth and is in turn controlled by it, thus preventing scarcity as well as waste, and thus combining the good features of competition with those of cooperation.

3. QUALITATIVE CHANGE MODELS

3.1. *Kinematical: Birth and Death Operators*

All sciences deal with qualitative changes and even deep ones such as changes in kind, or transmutations of things of one kind into things of another. Such changes in kind can be conceptualized in several ways. The simplest is by treating each kind as a graph node, and the transformation as a graph edge, of a directed graph. The latter is then a directed and connected graph – e.g. one representing a lineage such as a disintegration schema or a phylogenetic tree. If a quantitative, yet still purely phenomenological, representation of such qualitative changes is wanted, then one has to devise a way of saying how many things of each kind emerge or disappear in a given process, without suggesting any transmutation mechanisms. We shall exhibit one such model, inspired in quantum field theory.

Consider a system composed of things of different species (natural kinds) K_m, where m is a natural number greater than 2. Call n_m the instantaneous population of kind K_m in the system. The state of the system such that there are n_1 things of kind K_1, n_2 of kind K_2, and in general n_m of kind K_m, can be represented by a state function $u(n_1 n_2 \ldots n_i \ldots)$. Thus

$u(0000\ldots)$ represents the no-system state,

$u(1000\ldots)$ represents the state such that the system has a single component of kind K_1 and none of the others;

$u(0n0\ldots)$ represents the state of the system when it is composed of n units of kind K_2.

We assume that the totality of such state function values spans the state space S of the system. Moreover we postulate that S is a vector space, so that every linear combination of u's represents a possible state of the system. The unit vectors $\varepsilon_1 = (100\ldots)$, $\varepsilon_2 = (0100\ldots)$, $\varepsilon_3 = (00100\ldots)$ constitute a basis of the vector space.

We now introduce the *birth operators* $B_i \colon S \to S$ defined by

$$B_i u(n_1 \ldots n_i \ldots) = (n_i + 1)^{1/2} \, u(n_1 \ldots n_i + 1 \ldots).$$

B_i represents the change of state of the system brought about by the increase by one thing of kind K_i in the composition of the system.

Similarly we define the disappearance of one unit of kind K_i from the composition of the system:

$$D_i u(n_1 \ldots n_i \ldots) = (n_i)^{1/2} \, u(n_1 \ldots n_i - 1 \ldots),$$

and call $D_i \colon S \to S$ the *destruction operator* of kind K_i. It can be shown (e.g. Mandl, 1959) that the two operators are mutually adjoint, i.e. $D_i = B_i^\dagger$, where the dagger designates transposition and conjugation of the entries in the matrix representations of the operators. It follows that $B_i \, D_i$ leaves things in the state they were to begin with, i.e.

$$B_i D_i u(n_1 \ldots n_i \ldots) = n_i u(n_1 \ldots n_i \ldots).$$

The eigenvalues of $B_i D_i = B_i B_i^\dagger$ are the possible populations of kind K_i in the system. Therefore $B_i D_i = N_i$ is an operator that can be interpreted as the population (or occupation number) operator for the kind K_i.

The basic birth operators can be combined to yield operators representing arbitrary changes in the composition of the system. Thus $(B_i^\dagger)^p B_j{}^q$

represents the destruction of p individuals of kind K_i and the birth of q individuals of kind K_j. In other words,

$$(B_i^{\dagger})^p B_j^q \text{ represents the reaction } pK_i \longrightarrow qK_j,$$

whereas $(B_j^{\dagger})^p B_i^q$ represents the converse reaction. Mathematically speaking, $(B_i^{\dagger})^p B_j^q$ transforms the vector $u(n_1 \ldots n_i \ldots n_j \ldots)$ into $(n_i!(n_i + q)!/p!n_j!)^{1/2}$ $u(n_1 \ldots n_i - p \ldots n_j + q \ldots)$. Finally, the combinations of birth operators generate the entire state space out of the no-system (or zero) state $u(000\ldots)$ Thus $B_i^p B_j^q u(000\ldots)$ is proportional to the state vector representing the situation where there are p components of kind K_i and q of kind K_j in the system, and nothing else.

Because we have made no assumption other than that the changes under consideration are discrete changes in kind, the above is a framework for a kinematical (phenomenological) theory rather than a theory proper. But this is what we want in ontology: a more specific formalism would belong in science.

3.2. *Dynamical: Random Hits*

We shall presently sketch an extremely simple theory, both dynamical and stochastic, suggestive of a very general ontological hypothesis. The theory deals with the mutagenic effect of external agents impinging randomly upon a system. (See Maynard Smith (1968)). Suppose a system contains m components (*targets*) sensitive to certain discrete transmuting agents (*bullets*). The sensitive components must be a critical part of the whole thing even if they are only a small part: they may be atomic or cellular nuclei, organelles, key sociosystems, or what not. The disrupting agents can be ionizing particles, radiations, terrorists, bureaucrats, or what have you. Every hit or collision of a bullet with a target is assumed to either mutate or destroy the latter and, in any case, to effect a change in kind. The problem is to compute the probability that n random bullets will produce a change in kind (or transmutation) in a system with m sensitive components.

Assume that each bullet can act on a single target rather than being able to kill several birds at one stroke: i.e. let us restrict our attention to single hits. Call p the probability that an arbitrary bullet will mutate a target, and suppose that a total of n bullets fire at the system in an aimless and unconcerted way. Suppose also that the effect each successful bullet produces has no consequences on the other targets of the system – i.e. that the effects are mutually independent, so that only their cumulative effect is of interest. Since the probability that a bullet fails to hit a target is $q = 1 - p$, the probability that the target will fail to mutate under the action of n random bullets is

$$q^n = (1 - p)^n \cong e^{-pn} \quad \text{for large } n.$$

And the probability that all m sensitive components will remain unaltered after a round of n firings will be the m-fold product of the above, i.e.

$$Q \cong (e^{-pn})^m.$$

Consequently the probability that all m sensitive components will be hit and mutate is

$$P = 1 - Q \cong 1 - e^{-pmn},$$

which approaches 1 as the number of bullets increases. Therefore the average number of mutations induced by the n bullets is Pm, which approaches m with increasing n. In words: the probability that the bullets be ineffectual decreases exponentially as the number of bullets increases, while the probability that the system undergoes a radical transmutation increases monotonically with the number of bullets: see Figure B4.

Now, every system is subjected to some mutagenic agent or other – fast charged particles, slow neutrons, hard radiation, predators, critics, saboteurs, or what have you. Moreover, whether the missiles are aimed and concerted, or random, they are likely to have some effects – not necessarily destructive but in any case transmuting. Therefore we can make the sweeping generalization

POSTULATE B4 Every system except the world is subject to random actions that eventually elicit its breakdown or its mutation (change in kind).

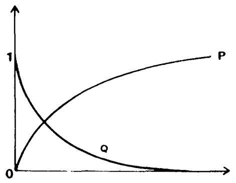


Fig. B4. Probable effects of random hits.

Note that we are not assuming that all mutagenic agents are external to the system concerned: it suffices that they be external to its sensitive components. For example, chromosomes are subjected not only to external radiation but also to thermal and chemical actions of the surrounding protoplasm. And every cultural system is the beneficiary or victim not only of economic and political actions but also of its own creative and destructive members.

BIBLIOGRAPHY

Alston, William P. (1974). Conceptual prolegomena to a psychological theory of intentional action. In Brown, Ed., 1974, pp. 71–101.

Anderson, James A. (1972). A simple neural network generating an interactive memory. *Math. Biosci.* **14**: 197–220.

Anderson, James A. (1973). A theory for the recognition of items from short memorized lists. *Psychol. Rev.* **80**: 417–438.

Arbib, Michael A. (1969). Automata theory. In Kalman *et al.*, 1969.

Apter, M. J., and Lewis Wolpert (1965). Cybernetics and development. *J. Theoret. Biol.* **8**: 244–257.

Aris, Rutherford (1965). Prolegomena to the rational analysis of systems of chemical equations. *Arch. Rational Mech. Anal.* **19**: 81–99.

Aristotle, *The Basic Works of A.* Richard McKeon, Ed. New York: Random House, 1941.

Armstrong, David M. (1968) *A Materialist Theory of the Mind.* London: Routledge & Kegan Paul.

Ashby, W. Ross (1956). *An Introduction to Cybernetics.* New York: Wiley.

Athans, M. and P. L. Falb (1966). *Optimal Control.* New York: McGraw-Hill.

Ayala, Francisco J. (1974). The concept of biological progress. In Ayala and Dobzhansky, Eds., 1974, pp. 339–355.

Ayala, Francisco J., and Theodosius Dobzhansky, Eds. (1974). *Studies in the Philosophy of Biology.* Los Angeles & Berkeley: University of California Press.

Bergson, Henri (1907). *L'évolution créatrice*, 77th ed. Paris: Presses Universitaries de France.

Berlinski, David (1976). *On Systems Analysis.* Cambridge, Mass.: MIT Press.

Bertalanffy, Ludwig von (1950). An outline of general system theory. *Brit. Jo. Philos. Sci.* **1**: 139–164.

Bertalanffy, Ludwig von (1968). *General System Theory.* New York: Braziller.

Bindra, Dalbir (1976). *A Theory of Intelligent Behavior.* New York: Wiley.

Birkhoff, Garrett, and Thomas C. Bartee (1970). *Modern Applied Algebra.* New York: McGraw-Hill Book Co.

Bogen, Joseph (1969). The other side of the brain: an appositional mind. *Bull. Los Angeles Neurol. Soc.* **34**, 3: 135–162.

Bonner, John Tyler (1974). *On Development. The Biology of Form.* Cambridge, Mass.: Harvard University Press.

Borst, C. V., Ed. (1970). *The Mind-Brain Identity Theory.* New York: St. Martin's Press.

Boulding, Kenneth (1956). General systems theory – the skeleton of science. *Management Sci.* **2**: 197–208.

Bowler, Peter J. (1976). *Fossils and Progress.* New York: Science History Publications.

Brandt, Richard, and Jaegwon Kim (1967). The logic of the identity theory. *J. Philos.* **64**: 515–537.

Braudel, Fernand (1949). *La Méditerranée et le monde méditerranéen à l'époque de Philippe II.* Paris: Armand Colin. Engl. transl.: *The Mediterranean.* New York: Harper and Row, 1973.

Braudel, Fernand (1969). *Écrits sur l'histoire.* Paris: Flammarion.

Brown, S. C., Ed. (1974). *Philosophy and Psychology.* London: Macmillan.

Bunge, Mario (1959a). *Causality: The Place of the Causal Principle in Modern Science.* Cambridge, Mass.: Harvard University Press, Rev. ed.: New York: Dover Publ., 1979.

Bunge, Mario (1959b). *Metascientific Queries.* Springfield, Ill.: Charles C. Thomas.

Bunge, Mario (1960). Levels: a semantical preliminary. *Rev. Metaphys.* **13**: 396–406. Repr. with slight changes in Bunge, 1963b.

Bunge, Mario (1963a). A general black box theory. *Philos. Sci.* **30**: 346–358.

Bunge, Mario (1963b). *The Myth of Simplicity.* Englewood Cliffs, N.J.: Prentice-Hall.

Bunge, Mario (1964). Phenomenological Theories. In M. Bunge, Ed., *The Critical Approach: Essays in Honor of Karl Popper.* Glencoe: The Free Press.

Bunge, Mario (1967a). *Scientific Research,* 2 vols. New York: Springer-Verlag.

Bunge, Mario (1967b). *Foundations of Physics.* New York: Springer-Verlag.

Bunge, Mario (1971). Is scientific metaphysics possible? *J. Philos.* 68: 507–520.

Bunge, Mario (1973a). *Method, Model, and Matter.* Dordrecht: D. Reidel Publ. Co.

Bunge, Mario (1973b). *Philosophy of Physics.* Dordrecht: D. Reidel Publ. Co.

Bunge, Mario (1974a). *Sense and Reference.* Dordrecht & Boston: D. Reidel Publ. Co.

Bunge, Mario (1974b). *Interpretation and Truth.* Dordrecht & Boston: D. Reidel Publ. Co.

Bunge, Mario (1974c). Les présupposés et les produits métaphysiques de la science et de la technique contemporaines. *Dialogue* **13**: 443–453.

Bunge, Mario (1974d). The concept of social structure. In W. Leinfellner and E. Köhler, Eds., *Developments in the Methodology of Social Science,* pp. 175–215. Dordrecht: D. Reidel Pub. Co.

Bunge, Mario (1976). A model for processes combining competition with cooperation. *Appl. Math. Model.* **1**: 21–23.

Bunge, Mario (1977a). *The Furniture of the World.* Dordrecht & Boston: D. Reidel Publ. Co.

Bunge, Mario (1977b). States and events. In William E. Hartnett, Ed., *Systems: Approaches, Theories, Applications.* Dordrecht & Boston: D. Reidel Publ. Co.

Bunge, Mario (1977c). The GST challenge to the classical philosophies of science. *Intern. J. General Systems* **4**: 29–37.

Bunge, Mario (1977d). General systems and holism. *General Systems* **22**: 87–90.

Bunge, Mario (1977e). Emergence and the mind. *Neurosci.* **2**: 501–509

Bunge, Mario (1977f). Levels and reduction. *Am. J. Physiol.* **R2**: 75–82.

Bunge, Mario (1977g). The philosophical richness of technology. In F. Suppe and P. D. Asquith, Eds., *PSA* **2**: 153–172.

Bunge, Mario (1977h) A systems concept of the international system. In *Bunge et al.,* Eds., pp. 291–305.

Bunge, Mario (1978a). A model of evolution. *Appl. Math. Model.* **2**: 201–204.

Bunge, Mario (1978b) A systems concept of society: beyond individualism and holism. Theory and Decision **10**: 13–30.

Bunge, Mario, and Máximo García Sucre (1976). Differentiation, participation and cohesion. *Quality and Quantity* **10**: 171–178.

Bunge, Mario, Johann Galtung, and Mircea Malitza, Eds. (1977) *Mathematical Approaches to International Relations.* Bucharest: Romanian Academy of Social and Political Sciences.

Bunge, Mario, and Rodolfo Llinás (1978). The mind-body problem in the light of neuroscience. *Proc. 16th World Congress of Philosophy.*

Calvin, Melvin (1969). *Chemical Evolution.* New York & Oxford: Oxford University Press.

Chiaraviglio, Lucio (1965). Sequential machines and the coding of polypeptides. *J. Theoret. Biol.* **8**: 130–140.

Chomsky, Noam (1968). *Language and Mind.* New York: Harcourt, Brace & World.

Cowan, Jack (1976). Are there modifiable synapses in the visual cortex? In Rosenzweig and Bennett, 1976, pp. 133–143.

Craik, Kenneth J. W. (1966), *The Nature of Psychology.* Stephen L. Sherwood, Ed. Cambridge: University Press.

Crick, Francis (1970), Central dogma of molecular biology. *Nature* **227**: 561–563.

Delgado, José M. R. (1969), *Physical Control of the Mind.* New York: Harper & Row.

Deutsch, Karl W. (1968), *The Analysis of International Relations.* Englewoods Cliffs, N. J.: Prentice-Hall.

Dobzhansky, Theososius (1970), *Genetics of the Evolutionary Process.* New York & London: Columbia University Press.

Dorf, Richard C. (1974), *Modern Control Systems,* 2nd ed. Reading, Mass.: Addison-Wesley Publ. Co.

Du Bois-Reymond, Emil (1872), Ueber die Grenzen des Naturerkennens. In *Reden,* 2nd ed. Leipzig: Veit und Co.

Dunbar, Max (1960). The evolution of stability in marine ecosystems: natural selection at the level of the ecosystem. *Am. Natur.* **94**: 129–136.

Edelen, Dominic G. B. (1962). *The Structure of Field Space.* Berkeley: University of California.

Eigen, Manfred (1971). Selforganization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**: 465–523.

Feigl, Herbert (1958), The 'mental' and the 'physical'. In H. Feigl, M. Scriven, and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* II, pp. 370–497. Minneapolis: University of Minnesota Press.

Feinberg, Martin (1972). On chemical kinetics of a certain class. *Arch. Rat. Mech. Anal.* **46**: 1–41.

Florkin, M. , Ed. (1960). *Aspects of the Origin of Life.* Oxford: Pergamon Press.

Fox, Sidney W., and K. Dose (1972). *Molecular Evolution and the Origin of Life.* San Francisco: Freeman & Co.

Freeman, Walter J. (1973). A model of the olfactory system. In M. A. M. Brazier, D. O. Walter, and D. Schneider, Eds., *Neural Modeling,* pp. 41–62. Los Angeles: UCLA Brain Research Institute.

Fritsch, Bruno (1974), *Wachstum Begrenzung als Machtinstrument.* Stuttgart: Deutsche Verlags-Anstalt.

Frondizi, Risieri (1953). *The Nature of the Self. A Functional Interpretation.* New Haven, Con.: Yale University Press. Repr.: Carbondale, Ill.: Southern Illinois Press, 1971.

Gall, John (1977) *Systemantics: How Systems Work and Especially How They Fail.* New York: Quadrangle.

Gardner, R. A., and B. T. Gardner (1969). Teaching sign language to a chimpanzee. *Science* **165**: 664–672.

Gazzaniga, Michael S. (1967). The split brain in man. *Sci. Am.* **217**, 2: 24–29.

Gierer, H., and H. Meinhardt (1972). A theory of biological pattern formation. *Kybernetik* **12**: 30–39.

Ginsburg, S. (1962). *An Introduction to Mathematical Machine Theory.* Reading, Mass.: Addison-Wesley Publ. Co.

Glansdorff, P., and I. Prigogine (1971). *Thermodynamic Theory of Structure, Stability and Fluctuations.* London: Wiley-Interscience.

Glass, Leon (1973). Stochastic generation of regular distributions. *Science* **180**: 1061–1063.

Goldin-Meadow, Susan, and Heidi Feldman (1977). The development of language-like communication without a language model *Science* **197**: 401–403.

Gray, J. A. (1972a). The psychophysiological nature of introversion-extroversion: A modification of Eysenck's theory. In Nebylitsyn and Gray, 1972, pp. 182–205.

Gray, J. A. (1972b). Learning theory, the conceptual nervous system, and personality. In Nebylitsyn and Gray, 1972, pp. 372–399.

Gray Eaton, G. (1976). The social order of Japanese macaques. *Sci. Am.* **235**, 4: 96–106.

Griffin, Donald R. (1976). *The Question of Animal Awareness.* New York: Rockefeller University Press.

Gross, Charles G., and H. Philip Zeigler, Eds. (1969). *Readings in Physiological Psychology. Motivation.* New York: Harper & Row.

Guinzburg, A. (1968). Algebraic Theory of Automata. New York: Academic Press.

Haken, H. (1975). Cooperative phenomena in systems far from equilibrium and in nonphysical systems. *Rev. Mod. Phys.* **47**: 67–121.

Hamilton, W. D. (1964). The genetical theory of social behavior. *J. Theoret. Biol.* **7**: 1–16, 17–32.

Hamilton, W. D. (1971), Geometry of the selfish herd. *J. Theoret. Biol.* **31**: 295–311.

Harris, Marvin (1968). *The Rise of Anthropological Theory.* New York: Thomas Y. Crowell Co.

Harrison, M. A. (1965). *Introduction to Switching and Automata Theory.* New York: McGraw-Hill.

Hebb, Donald O. (1949). *The Organization of Behavior.* New York: Wiley.

Hebb, Donald O. (1968). *A Textbood of Psychology.* Philadelphia: W. B. Saunders Co.

Hebb, Donald O. (1974).What psychology is about. *Am. Psychol.* **29**: 71–79.

Hegel, Georg Wilhelm Friedrich (1830). *Encyklopädie der philosophischen Wissenschaften im Grundrisse,* 3rd ed. Heidelberg: Verwaltung des Oswaldschen Verlags.

Henderson, Lawrence J. (1913). *The Fitness of the Environment.* Boston: Beacon Press, 1958.

Hinde, Robert A. (1974). *Biological Bases of Human Social Behaviour.* New York: McGraw-Hill.

Holland, John H. (1975). *Adaptation in Natural and Artificial Systems.* Ann Arbor: University of Michigan Press.

Homans, George C. (1974). *Social Behavior: Its Elementary Forms.* New York: Harcourt, Brace, Jovanovich.

Hubel, D. H., and T. N. Wiesel (1963), Receptive fields and functional architecture of monkey striate cortex. *J. Physiol.* **165**: 559–568.

Hull, David (1974), *Philosophy of Biological Science.* Englewood Cliffs, N.J.: Prentice-Hall.

Hume, David (1739). *A Treatise of Human Nature.* L. A. Selby-Bigge, Ed. Oxford: Clarendon Press, 1888.

Jacob, François (1970). *La logique du vivant.* Paris, Gallimard.

Jacob, François (1977). Evolution and tinkering. *Science* **196**: 1161–1166.

Jost, R. (1965). *The General Theory of Quantized Fields.* Providence: American Mathematical Society.

Kalman, R. E., P. L. Falb, and M. A. Arbib (1969). *Topics in Mathematical System Theory.* New York: McGraw-Hill.

Kawai, Masao (1965). Newly acquired pre-cultural behavior of the natural troop of Japanese monkeys on Koshima Islet. *Primates* 6: 1–30.

Kemeny, John, and Paul Oppenheim (1956). On reduction. *Philos. Stud.* **7**: 6–19.

Kendrew, John C. (1966). *The Thread of Life.* London: G. Bell and Sons Ltd.

Keosian, John (1968). *The Origin of Life,* 2nd ed. New York: Van Nostrand Reinhold Co.

King, M. -C., and A. C. Wilson (1975). Evolution at two levels in humans and chimpanzees. *Science* **188**: 107–116.

Klir, George (1969). *An Approach to General System Theory.* New York: Van Nostrand Reinhold Co.

Klir, Jiri, and Miroslav Valach (1967). *Cybernetic Modelling.* London: Iliffe Books Ltd.; Prague: SNTL.

Klir, George, and Gary Rogers (1977). *Basic and Applied General Systems Research: A Bibliography.* Binghampton, N.Y.: SUNY-Binghampton.

Kripke, Saul (1971). Identity and necessity. In Milton K. Munitz, Ed., *Identity and Individuation,* pp 135–164 New York: New York University Press.

Krohn, Kenneth, Rudolph Langer and John Rhodes (1967). Algebraic principles for the analysis of a biochemical system. *J. Comp. Syst. Sci.* **1**: 119–136.

Küppers, Bernd (1975). The general principles of selection and evolution at the molecular level. *Prog. Biophys. Molec. Biol.* **30**: 1–22.

Lange, Oscar (1965). *Wholes and Parts: A General Theory of Systems Behavior.* Oxford: Pergamon Press; Warsaw: PWN.

Lehninger, Albert L. (1975). *Biochemistry,* 2nd ed. New York: Worth Publishers Inc.

Lepper, M. R., D. Greene, and R. E. Nisbett (1973). Undermining children's intrinsic interest with extrinsic rewards. *J. Pers. Soc. Psychol.* **28**: 129–137.

Levins, Richard (1968). *Evolution in Changing Environments.* Princeton: Princeton University Press.

Lévi-Strauss, Claude (1963). *Structural Anthropology.* New York: Basic Books.

Lewontin, Richard C. (1974). *The Genetic Basis of Evolutionary Change.* New York & London: Columbia University Press.

Llinas, Rodolfo, and Mario Bunge (1978). Restricted applicability of the concept of command in neuroscience. *Behav. Brain Sci.* **1**: 30–31.

Lloyd Morgan, C. (1933). *The Emergence of Novelty.* London: Williams and Norgate.

Luria, A. R. (1966) *Human Brain and Psychological Processes.* New York: Harper & Row.

Luria, A. R. (1969) Speech development and the formation of mental processes. In Michael Cole and Irving Maltzman, Eds., *A Handbook of Contemporary Soviet Psychology*, pp. 121–162. New York: Basic Books.

Lwoff, André (1962). *Biological Order*. Cambridge, Mass.: MIT Press.

MacGregor, Ronald J., and Edwin R. Lewis (1977). *Neural Modelling*. New York & London: Plenum Press.

Malcolm, Norman (1973). Thoughtless brutes. *Proc. Add. Am. Philos. Assoc.* **46**: 5–20.

Malcolm, Norman (1964). Scientific materialism and the identity theory. *Dialogue* **3**: 115–125.

Mandl, F. (1959). *Introduction to Quantum Field Theory*. New York: Interscience.

Malsburg, C. von der (1973). Self-organization of orientation sensitive cells in the striate cortex. *Kybernetik* **14**: 85–100.

Margolis, Joseph (1978) *Persons and Minds: The Prospects of Nonreductive Materialism*. Dordrecht & Boston: D. Reidel Publ. Co.

Marr, David (1970). A theory for cerebral cortex. *Proc. Roy. Soc.* (London.) **B176**: 161–234.

Maynard Smith, J. (1968). *Mathematical Ideas in Biology*. Cambridge: University Press.

McNaughton, Robert (1961). The theory of automata: a survey. *Adv. Comp.* **2**: 379–421.

Menzel, E.W., Jr., R. K. Davenport, and C. M. Rogers (1972). Protocultural aspects of chimpanzee's responsiveness to novel objects. *Folia primatol.* **17**: 161–170.

Mesarović, M. D., Ed. (1964). *Views on General Systems Theory*. New York: Wiley.

Miller, Stanley, and Leslie Orgel (1974). *The Origins of Life on the Earth*. Englewood Cliffs, N.J.: Prentice-Hall.

Miller, Neal E. (1969). Learning of visceral and glandular responses. *Science* **163**: 434–445.

Milsum, John H., Ed. (1968). *Positive Feedback*. Oxford: Pergamon Press.

Monod, Jacques (1970). *Le hasard et la nécessité*. Paris: Ed. du seuil. Engl. transl.: *Chance and Necessity*. New York: Knopf, 1971.

Moore, Robert Y. (1976). Synaptogenesis and the morphology of learning and memory. In Rosenzweig and Bennet, Eds., pp. 340–347.

Morris, Charles W. (1938) Foundations of the theory of signs. *International Encyclopaedia of Unified Science*, vol. I., No. 2. Chicago: University of Chicago Press.

Mostow, G. D., Ed. (1975). *Mathematical Models for Cell Rearrangement*. New Haven, Conn.: Yale University Press.

Mountcastle, Vernon (1957). Modality and topographic properties of single neurons of cat's somatic sensory cortex. *J. Neurophys.* **20**: 408–434.

Nass, Menasche M., and Leon N. Cooper (1975). A theory for the development of feature detecting cells in visual cortex. *Biolog. Cybernet.* **19**: 1–18.

Nebylitsyn, V. D., and A. J. Gray, Eds. (1972). *Biological Bases of Individual Behaviour*. New York & London: Academic Press.

Neumann, John von (1951). The general and logical theory of automata. In L. A. Jeffress, Ed., *Cerebral Mechanisms in Behavior*. New York: Wiley.

Neumann, John von (1966). *Theory of Self-Reproducing Automata*. Ed. and completed by A. W. Burks. Urbana: University of Illinois Press.

Nicolis, G., and I. Prigogine (1977). *Self-Organization in Nonequilibrium Systems.* New York: Wiley-Interscience.

O'Connor, John, Ed. (1969). *Modern Materialism: Readings on Mind-Body Identity.* New York: Harcourt, Brace & World.

Olby, Robert (1974). *The Path to the Double Helix.* Seattle: University of Washington Press.

Omenn, Gilbert S., and Arno Motulsky (1972). Biochemical genetics and the evolution of human behavior. In L. Ehrman, G. S. Omenn, and E. Caspari, Eds., *Genetics, Environment and Behavior*, pp. 129–172. New York: Academic Press.

Oparin, A. I. (1968). *Genesis and Evolutionary Development of Life.* New York: Academic Press.

Oppenheim, Paul, and Hilary Putnam (1958). Unity of science as a working hypothesis. In H. Feigl, G. Maxwell, and M. Scriven, Eds., *Minnesota Studies in the Philosophy of Science*, Vol. II, pp. 3–36. Minneapolis: University of Minnesota Press.

Oró, J., S. L. Miller, C. Ponnamperuma, and R. S. Young, Eds. (1974). *Cosmochemical Evolution and the Origins of Life*, I. Dordrecht & Boston: D. Reidel Publ. Co.

Padulo, Louis, and Michael A. Arbib (1974). *System Theory.* Philadelphia: W. B. Saunders Co.

Paivio, A. (1971). *Imagery and Verbal Processes.* New York: Holt, Rinehart and Winston.

Penfield, Wilder (1975). *The Mystery of the Mind: A Critical Study of Consciousness and the Human Brain.* Princeton: University Press.

Penfield, Wilder, and T. Rasmussen (1950). *The Cerebral Cortex of Man.* New York: Macmillan.

Pirie, N. W. (1960). Chemical diversity and the origins of life. In Florkin, Ed., 1960, pp. 55–62.

Polanyi, Michael (1968). Life's irreducible structure. *Science* **160**: 1308–1312.

Popper, Karl R. (1968). Epistemology without a knowing subject. In B. van Rootselaar and J. F. Staal, Eds., *Logic, Methodology and Philosophy of Science* III, pp. 333–373. Amsterdam: North-Holland.

Popper, Karl R. (1972). *Objective Knowledge. An Evolutionary Approach.* Oxford: Clarendon Press.

Popper, Karl R. (1974). Autobiography. In P.A. Schilpp, Ed., *The Philosophy of Karl R. Popper*, I, pp. 1–181. La Salle, Ill.: Open Court.

Popper, Karl R., and John C. Eccles (1977). *The Self and its Brain.* Berlin, Heidelberg, London, New York: Springer International.

Powell, T.P.S., and Vernon Mountcastle (1959). Some aspects of the functional organization of the cortex of the postcentral gyrus of the monkey: a correlation of findings obtained in a single unit analysis with cytoarchitecture. *Bull. Johns Hopkins Hosp.* **105**: 133.

Premack, David (1971). Language in chimpanzee? *Science* **172**: 808–822.

Puccetti, Roland (1977). Sperry on consciousness: A critical appreciation. *J. Med. Philos.* **2**: 127–144.

Putnam, Hilary (1969). On properties. In Nicholas Rescher, Ed., *Essays in Honor of Carl G. Hempel*, pp. 235–254. Dordrecht: D. Reidel Publ. Co.

Rabin, Michael O. (1963). Probabilistic automata. *Information and Control* **6**: 230–245.

Rabin, Michael O., and Dana Scott (1959). Finite automata and their decision problems. *IBM J. Res. & Devel.* **3**: 114–125.

Rashevsky, Nicholas (1972). Some remarks on the central nervous system. *Bull. Math. Biophys.* **34**: 231–242.

Ratcliffe-Brown, A. R. (1935). On the concept of function in social science. *Am. Anthr.* **37**: 397–402.

Reynolds, Vernon (1976). *The Biology of Human Action.* Reading & San Francisco: W. H. Freeman and Co. Ltd.

Rosen, Robert (1959). The DNA-protein coding problem. *Bull. Math. Biophys.* **21**: 71–95.

Rosen, Robert (1970). Subunits and subassembly processes. *J. Theoret. Biol.* **28**: 415–422.

Rosenblueth, Arturo, Norbert Wiener, and Julian Bigelow (1943). Behavior, purpose and teleology. *Philos. Sci.* **10**: 18–24.

Rosenblueth, Arturo, and Norbert Wiener (1950). Purposeful and non-purposeful behavior. *Philos. Sci.* **17**: 318–326.

Rosenzweig, Mark R., and Edward L. Bennett, Eds. (1976). *Neural Mechanisms of Learning and Memory.* Cambridge, Mass.: MIT Press.

Rumbaugh, Duane M., Timothy V. Gill, and Ernst von Glasersfeld (1973). Reading and sentence completion by a chimpanzee (*Pan*). *Science* **182**: 731–733.

Ruse, Michael (1979). *Sociobiology: Sense or Nonsense*? Dordrecht & Boston: D. Reidel Publ. Co.

Rutledge, L. T. (1976). Synaptogenesis: effects of synaptic use. In Rosenzweig and Bennett, Eds., 1976, pp. 329–339.

Schlick, Moritz (1925). *Allgemeine Erkenntnislehre*, rev. ed. Transl. A. E. Blumberg, *General Theory of Knowledge.* Wien-New York: Springer-Verlag, 1974.

Sellars, Roy Wood (1922). *Evolutionary Naturalism.* Chicago: Open Court.

Sellers, Peter H. (1967). Algebraic complexes which characterize chemical networks. SIAM *J. Appl. Math.* **15**: 13–68.

Sellers, Peter H. (1970). Combinatorial analysis of a chemical network. *J. Franklin Inst.* **290**: 113–130.

Sellers, Peter H. (1971). An introduction to a mathematical theory of chemical reaction networks I. *Arch. Rational Mech. Anal.* **44**: 22–40.

Shaffer, Jerome A. (1977). Personal identity: the implications of brain bisection and brain transplant. *J. Medicine and Philosophy* **2**: 147–161.

Smart, J. J. C. (1959). Sensations and brain processes. *Philos. Rev.* **68**: 141–156.

Smith, Emil (1970). Evolution of enzymes. In Paul D. Boyer, Ed., *The Enzymes*, Vol. I, pp. 267–339. New York: Academic Press.

Smith, W. John (1965). Message, meaning, and context in ethology. *Am. Natur.* **99**: 405–409.

Smythies, J. R., Ed. (1965). *Brain and Mind.* London: Routledge & Kegan Paul.

Sommerhoff, Gerd (1974). *Logic of the Living Brain.* New York: Wiley.

Spencer, Herbert (1862). *First Principles.* Final ed.: London: Watts & Co., 1937.

Sourkes, Theodore (1962). *Biochemistry of Mental Disease.* New York: Harper & Row.

Sperry, Roger W. (1964). Neurology and the mind-body problem. In Robert Isaacson, Ed., *Basic Readings in Neuropsychology*, pp. 403–429. New York: Harper & Row.

Sperry, Roger W. (1966). Brain bisection and mechanisms of consciousness. In John C. Eccles, Ed., *Brain and Conscious Experience.* Berlin, Heidelberg & New York: Springer-Verlag, 1966.

Sperry, Roger W. (1977). Forebrain commissurotomy and conscious awareness. *J. Med. Philos.* **2**: 101–126.

Spinoza, Benedict (1677). *Ethics.* Oxford: University Press, 1927.

Steinberg, Malcolm S. (1963). Reconstruction of tissues by dissociated cells. *Science* **141**: 401–408.

Thom, René (1972). *Stabilité structurelle et morphogenèse.* Reading, Mass.: W. A. Benjamin Inc.

Toulmin, Stephen (1972). The mentality of man's brain. In A.G. Karczmar and John C. Eccles, Eds., *Brain and Human Behavior*, pp. 409–422. New York, Heidelberg & Berlin: Springer-Verlag.

Turing, Alan (1950). Computing machinery and intelligence. *Mind* NS**59**: 433–460.

Turing, Alan (1952). The chemical basis of morphogenesis. *Philos. Trans. Roy. Soc.* (*Lond.*) **B237**: 7–94.

Tyson, John J. (1976). *The Belousov–Zhabotinskii Reaction.* Berlin, Heidelberg, New York: Springer-Verlag.

Vaucouleurs, Gérard de (1970). The case for a hierarchical cosmology. *Science* **167**: 1203–1213.

Vesey, G. N. A., Ed. (1964. *Body and Mind.* London: George Allen & Unwin.

Vol'kenshtein, Mikhail V. (1970). *Molecules and Life.* New York: Plenum Press.

Vygotskii, L. S. (1962). *Thought and Language.* Transl. E. Hanfman and G. Vakar. Cambridge, Mass.: MIT Press.

Waddington, Conrad Hall (1959). Evolutionary systems – animal and human. *Nature* **183**: 1654–1658.

Wallerstein, Immanuel (1974). *The Modern World-System: Capitalist Agriculture and the Origins of the European World Economy in the Sixteenth Century.* New York: Academic Press.

Watson, James (1976). *Molecular Biology of the Gene*, 3rd ed. Menlo Park, Cal.: W. A. Benjamin, Inc.

Wiener, Norbert (1948). *Cybernetics.* New York: John Wiley & Sons.

Williams, George C. (1966). *Adaptation and Natural Selection.* Princeton: Princeton University Press.

Wilson, Edward O. (1975). *Sociobiology: The New Synthesis.* Cambridge, Mass.: Harvard University Press.

Wilson, H. R., and J. D. Cowan (1973). A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. *Kybernetic* **13**: 55–80.

Winch, Peter (1958). *The Idea of a Social Science.* London: Routledge & Kegan Paul.

Worden, Frederic G., Judith P. Swazey, and George Adelman, Eds. (1975). *The Neurosciences: Paths of Discovery.* Cambridge, Mass.: MIT Press.

Zadeh, Lofti A., and Charles A. Desoer (1963). *Linear System Theory: The State Space Approach.* New York: McGraw-Hill.

INDEX OF NAMES

Adelman, George 124
Alexander, Samuel 250
Alston, William P. 144
Anderson, James A. 162
Apter, M. J. 116
Arbib, Michael A. 253, 263
Aris, Rutherford 59
Aristotle 102, 117, 216, 219
Armstrong, David M. 126, 141
Ashby, W. Ross 253
Athans, M. 253
Axelrad, D. R. xiii
Ayala, Francisco J. 116

Bartee, Thomas C. 71
Bergson, Henri 102, 103
Berlinski, David 16
Bertalanffy, Ludwig von 1
Berzelius 48
Bettuci, Lina xiii
Bigelow, Julian 164
Bindra, Dalbir 128, 131, 140, 152, 162
Birkhoff, Garrett 71
Blohm, Robert xiii
Bogen, Joseph 142
Bois-Reymond, E. du 124
Bonner, John Tyler 95
Borst, C. V. 126
Boulding, Kenneth 1
Bowler, Peter J. 116
Brandt, Richard 140
Braudel, Fernand 207, 240
Bravo, Ernesto M. xiii
Brito da Cunha, A. xiii
Bunge, Mario 1, 3, 4, 13, 14, 22, 41, 42, 86, 108, 119, 122, 126, 143, 148, 207, 208, 209, 213, 225, 226, 229, 231, 237, 243, 255, 257, 279, 283, 285, 286

Calvin, Melvin 28
Chiaraviglio, Lucio 52
Chomsky, Noam 158
Cooper, Leon N. 137
Cowan, Jack 133, 137
Craik, Kenneth J. W. 128
Crick, Francis xiii, 68

Darwin, Charles 29, 104, 107
Davenport, R. K. 182
Delgado, José M. R. 142
Descartes, René 120, 141
Desoer, Charles A. 253
Deutsch, Karl W. 207
Dobzhansky, Theodosius 83, 103, 106
Dorf, Richard C. 253
Dose, K. 78
Dubrovsky, Bernardo xiii
Dunbar, Max 108

Eccles, John C. 4, 124, 126, 141, 148, 170, 171, 175
Edelen, Dominic 279
Eigen, Manfred 28, 60, 92
Epicurus 88, 144

Fabi, Ricardo Peralta xiii
Falb, P. L. 253
Fechner, 142
Feigl, Herbert 147
Feinberg, Martin 59
Feldman, Heidi 181
Florkin, M. 78
Fox, Sidney W. 78
Freeman, Walter J. 131
Freud, Sigmund 172, 178
Fritsch, Bruno 207

Gall, John 3

Galtung, Johann 207
García de la Sienra, Adolfo xiii
García-Sucre, Máximo 229, 231
Gardner, B. T. 181
Gardner, R. A. 181
Gazzaniga, Michael S. 142
Gierer, H. 96
Gill, Timothy V. 181
Ginsburg, S. 263
Glansdorff, P. 96, 248
Glasersfeld, Ernst von 181
Glass, Leon 96
Goethe, Johann Wolfgang von 250
Goldin-Meadow, Susan 181
Graham, Gordon xiii
Gray, J. A. 147, 174
Gray Eaton, G. 182
Greene, D. 214
Grey Walter, W. 121
Griffin, Donald R. 172
Gross, Charles G. 159
Guinzberg, A. 263

Haken, H. 248
Hamilton, W. D. 94, 178, 179
Harris, Marvin 234
Harrison, M. A. 253, 263
Hawthorne, John xiii
Hebb, Donald O. 130, 133, 147, 151–152, 173
Hegel, Georg Wilhelm Friedrich 48
Henderson, Lawrence J. 76
Heraclitus 284
Hinde, Robert A. 175
Hippocrates 124
Hitschfeld, Walter xiii
Holland, John H. 105
Homans, George C. 234
Hubel, D. H. 131, 150
Hull, David 82
Hume, David 174

Jacob, François 76, 92, 119
Jost, R. 280

Kawai, Masao 182
Kemeny, John 43
Kendrew, John C. 119
Keosian, John 119
Kim, Jaegwon 140
King, M.-C. 122
Klir, George 15, 16, 253
Kripke, Saul 148
Krohn, Kenneth 59
Küppers, Bernd xiii, 73

Lagrange, Joseph-Louis 280
Lamarck, Jean-Baptitse de 107, 117
La Mettrie, Julien Offray de 120
Lange, Oscar 253
Langer, Rudolph 59
Lehninger, Albert L. 28, 60, 62, 63
Leibniz, Gottfried Wilhelm von 102
Lepper, M. R. 214
Levins, Richard 106
Lévi-Strauss, Claude 234
Lewis, Edwin R. 131
Lewontin, Richard C. 108
Llinás, Rodolfo xiii, 178
Lloyd Morgan, Conwy 103
Lorenz, Konrad 178
Luria, A. R. 131, 175
Lwoff, André 116

Malcolm, Norman 148, 181
Malitza, Mircea 207
Malsburg, C. von der 133
Margeneau, Henri 280
Margolis, Joseph 126, 184
Marr, David 137
Marx, Karl 216, 219
Maupertuis, Pierre Louis 280
Maxwell, James Clerk 280
Maynard Smith, John 289
MacGregor, Ronald J. 131
McNaughton, Robert 263
Meinhardt, H. 96
Menzel, E. W. 182
Mesarović, M. D. 254
Miller, Neal E. 142
Miller, Stanley 76, 78
Monod, Jacques 92, 99, 119
Moore, Robert Y. 133
Morgan, Henry 139

Morris, Charles W. 180
Mostow, G. D. 97
Motulsky, Arno 143
Mountcastle, Vernon 131

Nass, Menasche M. 137
Neumann, John von 120, 264
Nicolis, G. 96
Nisbett, R. E. 214

O'Connor, John 126
Olby, Robert 121
Omenn, Gilbert S. 143
Oparin, A. I. 76, 77, 78
Oppenheim, Paul 43
Orgel, Leslie 76, 78
Oró, J. 76, 78

Padulo, Louis 253
Paivio, A. 180
Palfree, Roger xiii
Parmenides 153
Pascual, Rafael Pérez xiii
Penfield, Wilder 145
Piaget, Jean 175
Pirie, N. W. 119
Planck, Max 280
Plato 170
Polanyi, Michael 120
Ponnamperuma, C. 76, 78
Popper, Karl 4, 124, 126, 141, 148, 170, 171, 175, 234
Powell, T. P. S. 131
Premack, David 181
Prigogine, I. 96, 248
Puccetti, Roland 143
Putnam, Hilary 43

Rabin, Michael O. 263
Ramón y Cajal 124, 133
Rashevsky, Nicholas 131
Rasmussen, T. 145
Ratcliffe-Brown, A. R. 10
Reig, Osvaldo A. xiii
Rhodes, John 59
Rogers, C. M. 182
Rogers, Gary 15
Rosen, Robert 52, 249
Rosenblueth, Arturo 164
Rumbaugh, Duane M. 181
Ruse, Michael 179
Rutledge, L. T. 133

Schlick, Moritz 147
Scott, Dana 263
Sellars, Roy Wood 251
Sellers, Peter H. 59
Seni, Daniel xiii
Shaffer, Jerome 174
Shannon, Claude 271
Smart, J. J. C. 148
Smith, W. John 180
Smythies, J. R. 126
Sommerhoff, Gerd 120
Spencer, Herbert 104, 124
Sperry, Roger W. 142, 171
Spinoza, Benedictus 158
Steinberg, Malcolm S. 97, 98
Swazy, Judith P. 124
Szent-György, Albert 62

Thom, René 97, 102
Tocqueville, Alexis de 219
Toulmin, Stephen 175
Turing, Alan 97, 248
Tyson, John J. 96

Valach, Miroslav 16, 253
Vaucouleurs, Gérard de 28
Vesey, G. N. A. 126
Vol'kenstein, Mikhail 62
Vygotskii, L. S. 175

Waddington, Conrad Hall 105–106
Wallerstein, Immanuel 207
Watson, James 62, 68
Weaver, Warren 271
Wiener, Norbert 164, 197, 258
Wiesel, T. N. 131, 150
Williams, George C. 104, 116
Wilson, A. C. 122
Wilson, Edward O. 179
Wilson, H. R. 137
Winch, Peter 234

Wittgenstein, Ludwig 181
Wolpert, Lewis 116
Worden, Frederic G. 124

Young, R. S. 76, 78

Zadeh, Lofti A. 253
Zayan, René xiii
Zeigler, H. Philip 159

INDEX OF SUBJECTS

action 6, 9, 24
adaptation 103–106
aggregate 3–4, 11, 21–22
altruism 178–179
amino acid 62–63, 65–66
animal 129–130
artifact 209–210
assembly 27–28, 30, 97–98
association 14
atom 46–51
atomism 41–44
automata 263–270

balance 275–277
behavior 156–159
  social 175–176, 179–180, 225
biofunction 89–91
biolevel 84–85
biomolecule 61
biopopulation 83, 85
biospecies 83
biosphere 83–84
birth and death operators 287–289
black box 253–257, 260
bond 6, 27, 35
  social 221–226, 231–233, 235
brain 126, 138–151, 174–175

catalysis 60
cell 82, 88–89
chance 76–77
change 22, Appendix B
  see also event, evolution, history, process
  social 235–241
chemical 51
choice 160–161
class, economic 228
  political 228
  social 228–229
  structure 228–229
closure *see* system, closed
codon 70–71
coevolution 107–108
cognition 166
cohesion *see* systemicity
  social 177, 231–233
communication 180–181
community *see* ecosystem
composition (of a system) 4–8, 11, 43–44
compound 51
concept 165
connection 6, 9
connectivity 131–132
consciousness 170–172
control 59–61
  biocontrol 92–95
  enzymatic 92–93
  genetic 92–93, 95
cooperation 81, 177–178, 283–287
coupling graphs 16–17, 19
creativity 167–169, 182–183
culture 182–183, 204, 211–215

Darwinism 81
decision 167
death 87–88, 94
descent 33–34
detection 150
development 95, 101–102
differentiation 95–96, 98
DNA 62, 64–69
double helix 67
drive 159
dualism 124, 126–127, 129, 138–141, 143–144, 173, 183
  interactionist 141–142
dynamical model 281–283

economy 182–183, 204, 208–209, 214–215
ecosystem 83, 108
effect 24
emergence 29–30, 75, 184, 191, 249–251
environment 4–9, 32–33, 43–44
epigenesis 102–103
equity 217–219
event 22–23
evolution 33–34, 103–118, 186–187
  social 235–240
expectation 163–164
experience 162
extrasensory perception 139

family 176, 223–224
feedback 92–93, 164, 239, 257–260, 262
feeling *see* sensation
foresight *see* expectation
freedom 219–221
free will 173
function, biological 89–91
  mental 138–141

general systems theory 1
genetic code 70–71
genome 100–101
goal 91, 92, 94–95, 104–105, 118, 120, 164, 210–211
grey box 262–263
growth 95–96

Hamiltonian model 282–283
health 86–88, 90, 94
heredity 98–102
hierarchy 12–14, 84
history 240
holism 39–41, 43–44, 194, 234, 241–242, 250

idea 146, 169–170
identity theory 140–141, 148
information 9, 130, 270–272
institution 193–196
integration *see* systemicity
international relations 207–208
isomer 66

kinematics 273–275, 280–281
kinetic theory of gases 21
kinship 182
knowledge 167

Lagrangian model 277–280
language 180–181
learning 133, 162–164
level 13–14
life Chapter 3
  origin of 76–79
lineage 34
link *see* bond

materialism 126–127, 138, 146, 183–185
matrices 16–19
mechanism *see* metaphysics
memory 161–162
mental activity 138–142
mental process 138–141
mental state 138–141
metaphysics
  biosystemism 119–120, 122
  machinism 119–121
  mechanism 118–121
  physicochemicalism 119–122
  preformation vs. emergence 102–103
  substance & property 101–102
  vitalism 118–122
mind Chapter 4, *in particular* 139–147
  location 144–146
  mind-body problem 184
model, mathematical 2
molecule 46–52
mononucleotide 62–63
morphogenesis 95–98
morphology 96–97
mutation 99–100

nation 205–207
naturalism *see* materialism
natural kind 82–83
nervous system 130, 157

  autonomic 144
  central (CNS) 127, 129, 130, 134–136, 153
  neural system 130–133, 136–137, 140, 150, 157

ontology 3, 75, 251–252
optimality 94
ordinary language 124, 125, 148–149
organism 82

perception 151–155
person 173–174
phenome 100–101
phenomena 147
physics 1
Platonism 101, 118
polity 182–183, 204, 214–221
polymer 51–52
polypeptide 66
population 239
process 22–24
progress 115–118
proposition 165–166
protein 62–63, 65–67
psychon 128, 132
purpose *see* goal

rationality 167
reaction 54–60
reduction 42, 75
reproduction 98–99
RNA 62, 64–70
rules 196

selection 30–32, 107
self 175
sensation 151
sickness *see* health
social change *see* evolution, social
society 177, 182–183, Chapter 5
solidarity 178
spontaneous activity 25
state function 20–21, 23
state space 20–23
structure 4–8, 10–13, 43–44
  social 226–235
subsystem 11
system
  biochemical 61, 69, 71–74, 76, 85
  biosystem 79–88, 246–247
  chemical 52–55, 59–61, 69, 72–74, 246–247
  closed 9–10
  conceptual 4
  concrete 4–6, 8, 11, 25, 45
  coordination 38–39
  functional 15, 139–140
  inhibition 38–39
  input/output 24–26, 253–257, 261
  open 17–18
  physical 54, 61, 73–74, 246–247
  precursor 29
  property 15
  size of 36–37
  social 246
  systems analysis 2
  technical 246–247
systemicity 35–38
systemics 1–3

template 67–69
thing 14
thinking 164–166

universe 9–10, 245–246

value 89–90, 116–117, 159–160
vitalism *see* metaphysics
volition 172–173

wholes 39–41
will *see* volition
work 196–204
world view Chapter 6

CONTENTS OF *ONTOLOGY I*

GENERAL PREFACE TO THE *TREATISE* V

PREFACE TO *ONTOLOGY I* XIII

ACKNOWLEDGEMENTS XV

SPECIAL SYMBOLS XVI

INTRODUCTION 1
1. Ontological Problems 1
2. The Business of Ontology 3
3. Is Ontology Possible? 6
4. The Method of Scientific Ontology 7
5. The Goals of Scientific Ontology 9
6. Ontology and Formal Science 13
7. The Ontology of Science 15
8. Ontological Inputs and Outputs of Science and Technology 19
9. Uses of Ontology 22
10. Concluding Remarks 24

1. SUBSTANCE 26
1. Association 26
    1.1. Concatenation and its Ontological Interpretation 26
    1.2. Axiomatic Foundation of Association Theory 28
    1.3. Consequences 31
    1.4. Atom Aggregates 36
    1.5. Clustering 37
    1.6. Historical Remark 38

2. Assembly 39
    2.1. Juxtaposition and Superposition: Intuitive Idea 39
    2.2. Formalization 40
    2.3. Definitions 42
    2.4. Some Consequences 45

2.5. Atoms and Levels 47
2.6. Alternative Formalizations 49
2.7. Concluding Remarks 50

3. Entities and Sets 50
3.1. The Null Individual and the World 50
3.2. Entities and Concepts 52
3.3. Existence and Individuation 53

4. Concluding Remarks 55

2. FORM 57
1. Property and Attribute 58
1.1. Difference between Property and Attribute 58
1.2. Attribute-Property Correspondence 59

2. Analysis 62
2.1. Property in General and Property of a Particular 62
2.2. Intrinsic and Mutual, Primary and Secondary 65

3. Theory 69
3.1. Unarization and Dichotomization 69
3.2. Basic Assumptions and Conventions 72
3.3. Laws as Properties 77
3.4. Precedence and Conjunction of Properties 80
3.5. Similarity 85
3.6. Indiscernibility 90

4. Properties of Properties 92
4.1. Identity and Difference of Properties 92
4.2. Property Weight 94
4.3. Resultants and Emergents 97
4.4. Properties of Properties 98

5. Status of Properties 99
5.1. The Reality of Properties 99
5.2. A Critique of Platonism 102
5.3. The Problem of Universals 104

6. Concluding Remarks 108

3. THING 110
1. Thing and Model Thing 110
1.1. Thing: Definition 110

1.2. Assumptions 112
1.3. Thing and Construct 116
1.4. Model Thing 119

2. State 123
   2.1. Centrality of the State Concept 123
   2.2. State Function 125
   2.3. Law Statements as Restrictions on State Functions 128
   2.4. State Space: Preliminaries 131
   2.5. Definition of a State Space 133
   2.6. Equivalent Representations of States 136
   2.7. State and State Preparation 138
   2.8. Concluding Remarks 139

3. From Class to Natural Kind 140
   3.1. Classes of Things 140
   3.2. Ideals and Filters of Classes of Things 140
   3.3. Kinds and Species 143
   3.4. The Algebra of Kinds 147
   3.5. Variety 150

4. The World 152
   4.1. What Does the World Consist in and of? 152
   4.2. Individuals, Populations, Communities, and Species 153
   4.3. Existence Concepts 155
   4.4. Nothingness and Virtual Existence 158
   4.5. Existence Criteria 160

5. Concluding Remarks 162

4. POSSIBILITY 164
1. Conceptual Possibility 165
   1.1. Possibility Concepts 165
   1.2. Four Concepts of Conceptual Possibility 165
   1.3. Conceptual Possibility: Relative 168

2. Real Possibility 168
   2.1. Fact 168
   2.2. Chrysippian Possibility 172
   2.3. Real Possibility as Lawfulness 173
   2.4. Factual Necessity 174
   2.5. Possibility Criteria 177

3. Disposition 179
   3.1. Intuitive Idea 179

3.2. Elucidation 181
3.3. Potency and Act 183
3.4. Unrealized Possibilities and Counterfactuals 184

4. Probability 185
    4.1. Abstract Concept 185
    4.2. Probability State Space 187
    4.3. Propensity Interpretation 190

5. Chance Propensity 194
    5.1. Irreducible Potentialities 194
    5.2. Analysis 197
    5.3. Upshot 198

6. Marginalia 199
    6.1. Modal Logic and Real Possibility 199
    6.2. Possible Worlds Metaphysics 202
    6.3. Modality and Probability 204
    6.4. Randomness 208
    6.5. Probability and Causality 210
    6.6. The Many Worlds Interpretation of Quantum Mechanics 211

7. Concluding Remarks 212

5. CHANGE 215
1. Changeability 216
    1.1. Preliminaries 216
    1.2. Changeability 218

2. Event 221
    2.1. The Ordered Pair Representation of Events 221
    2.2. The Event Space 222
    2.3. The Representation of Processes 226
    2.4. The Space of Lawful Events 229
    2.5. Keeping Track of Changing States 232
    2.6. Rate, Extent, and Change Potential 237

3. Process 243
    3.1. Serial Change: Types 243
    3.2. General Concepts and Principles 251

4. Action and Reaction 256
    4.1. Induced Change 256
    4.2. Aggregates and Systems 262
    4.3. Reference Frame 264

5. *Panta Rhei* 267
  5.1. Fact 267
  5.2. Dynamicism 268
  5.3. Interconnectedness 270
  5.4. Three Misconceptions 271

6. Concluding Remarks 273

6. SPACETIME 276
1. Conflicting Views 278
  1.1. The Three Main Views 278
  1.2. Approaches to Chronotopics Building 281

2. Space 283
  2.1. Interposition 283
  2.2. A Philosopher's Space 285
  2.3. The Physicist's Space 287
  2.4. Bulk and Shape 293
  2.5. Concluding Remarks 294

3. Duration 296
  3.1. Intuitive Idea 296
  3.2. Before and After 297
  3.3. Duration 301

4. Spacetime 305
  4.1. Spacetime, the Basic Network of Events 305
  4.2. Position in Spacetime 309
  4.3. Change in Spacetime 312

5. Spatiotemporal Properties 314
  5.1. Does Spacetime have any Properties? 314
  5.2. Time Reversal and Process Reversibility 317
  5.3. Antecedence ("Causality") Principle 320
  5.4. Action by Contact 323
  5.5. Spatiotemporal Contiguity 325
  5.6. The Causal Relation 326

6. Matters of Existence 327
  6.1. Existence in Space and Time 327
  6.2. Existence of Space and Time 328

7. Concluding Remarks 330

BIBLIOGRAPHY 334

INDEX OF NAMES 344

INDEX OF SUBJECTS 348

TREATISE ON BASIC PHILOSOPHY

Volume 5

EPISTEMOLOGY AND METHODOLOGY I: EXPLORING THE WORLD

TREATISE ON BASIC PHILOSOPHY

1

SEMANTICS I *Sense and Reference*

2

SEMANTICS II *Interpretation and Truth*

3

ONTOLOGY I *The Furniture of the World*

4

ONTOLOGY II *A World of Systems*

5

EPISTEMOLOGY & METHODOLOGY I *Exploring the World*

6

EPISTEMOLOGY & METHODOLOGY II *Understanding the World*

7

EPISTEMOLOGY & METHODOLOGY III *Philosophy of Science & Technology*

8

ETHICS *The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 5

Epistemology & Methodology I:

EXPLORING THE WORLD

D. REIDEL PUBLISHING COMPANY
A MEMBER OF THE KLUWER ACADEMIC PUBLISHERS GROUP
DORDRECHT / BOSTON / LANCASTER

Library of Congress Cataloging in Publication Data

Bunge, Mario Augusto.
Exploring the world.

(Epistemology & methodology ; 1) (Treatise on basic philosophy ; v. 5)
Bibliography: p.
Includes indexes.
1. Knowledge, Theory of. 2. Methodology. I. Title. II. Series: Bunge, Mario Augusto. Epistemology and methodology, 1. III. Series: Bunge, Mario Augusto. Treatise on basic philosophy ; v. 5.
BD161.B86 1983 121 83-3236
ISBN-13: 978-90-277-1523-4 e-ISBN-13: 978-94-009-7027-4
DOI: 10.1007/978-94-009-7027-4

Published by D. Reidel Publishing Company,
P.O. Box 17, 3300 AA Dordrecht, Holland.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers
190 Old Derby Street, Hingham, MA 02043, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, Holland.

All Rights Reserved
© 1983 by D. Reidel Publishing Company, Dordrecht, Holland
Softcover reprint of the hardcover 1st edition 1983
No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any information storage and
retrieval system, without written permission from the copyright owner

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and of right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient—except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can

become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: "Do your own thing. Your reward will be doing it, your punishment having done it".

CONTENTS OF EPISTEMOLOGY I

GENERAL PREFACE TO THE *TREATISE* v

PREFACE TO *EPISTEMOLOGY I & II* xiii

ACKNOWLEDGEMENTS xvii

SPECIAL SYMBOLS xix

INTRODUCTION 1

1. Descriptive Epistemology 1
2. Normative Epistemology (Methodology) 4
3. Epistemology and Biology 6
4. Epistemology and Psychology 8
5. The Place of Epistemology 10
6. Uses of Epistemology 13
7. Concluding Remarks 16

PART 1. COGNITION AND COMMUNICATION 19

1. COGNITION 21
   1. The Knowing Subject 23
      - 1.1. The Cognitive Organ 23
      - 1.2. Brain States and Processes 28
   2. Cognitive Functions 35
      - 2.1. Perceiving, Thinking, and all that 35
      - 2.2. Learning 42

3. Development and Evolution 50
  3.1. Cognitive Development 50
  3.2. Cognitive Evolution 57

4. Concluding Remarks 59

2. KNOWLEDGE 61
1. From Cognition to Knowledge 62
  1.1. Knowledge Acquisition and Use 62
  1.2. Knowledge Evolution 67

2. Modes of Knowledge 72
  2.1. Basic Categories 72
  2.2. Tacit and Explicit Knowledge 77

3. Belief 80
  3.1. Belief System 80
  3.2. Belief and Knowledge 86

4. Inquiry 88
  4.1. Field of Knowledge 88
  4.2. How is Knowledge Possible? 92

5. Concluding Remarks 96

3. COMMUNICATION 97
1. Subject and World 97
  1.1. Subject–Object Relation 97
  1.2. Inter-Subject Relation 101

2. Social Matrix 103
  2.1. Cognitive Habitat and Niche 103
  2.2. Learning Community 108

3. Communication 112
  3.1. Communication System 112
  3.2. Language, Thought, and Reality 120

4. Concluding Remarks 124

PART II. PERCEIVING AND THINKING 127

4. PERCEIVING 129
  1. From Sensing to Perceiving 130
    1.1. Sensing 130
    1.2. Perceiving 133
  2. Observing 139
    2.1. Direct Observation 139
    2.2. Indirect Observation 142
  3. Mapping Reality 146
    3.1. Perceptual Mapping 146
    3.2. Appearance and Reality 150
  4. Concluding Remarks 155

5. CONCEIVING 159
  1. Concept 159
    1.1. From Percept to Concept 159
    1.2. Concept Formation 164
  2. Proposition 174
    2.1. From Thought to Proposition 174
    2.2. A Priori and a Posteriori 180
  3. Exactification 182
    3.1. From Vagueness to Exactness 182
    3.2. Quality and Quantity 187
  4. Representation 191
    4.1. Conceptual Map 191
    4.2. Discovery or Invention? 194
  5. Concluding Remarks 196

6. INFERRING 199
  1. From Spontaneity to Formality 200
    1.1. Natural Reasoning 200
    1.2. Formal Reasoning 203
  2. Plausible Reasoning 210
    2.1. Analogical Inference 210
    2.2. Statistical Inference and Hypothetical Reasoning 214

3. Discussing 218
  3.1. Analysis and Criticism 218
  3.2. Controversy 224

4. Concluding Remarks 228

PART III. EXPLORING AND THEORIZING 231

7. EXPLORING 233
  1. Exploratory Behavior 233
    1.1. Psychobiology of Problems 233
    1.2. Sources and Choices 241

  2. From Intuition to Method 248
    2.1. Intuition 248
    2.2. Method 250

  3. Approach and Research 258
    3.1. Approach 258
    3.2. Research 262

  4. Analysis of Problems 269
    4.1. Logic and Semantics of Problems 269
    4.2. Solved, Unsolved, Insoluble 280

  5. Concluding Remarks 284

8. CONJECTURING 286
  1. From Preconception to Hypothesis 286
    1.1. Psychobiology of Conjecturing 286
    1.2. Hypothesis 290

  2. Scope and Depth 297
    2.1. Scope 297
    2.2. Depth 300

  3. Chance and Cause 305
    3.1. Probabilistic Hypotheses 305
    3.2. Causal Hypotheses 310

4. Requirements 314
4.1. Strength and Precision 314
4.2. Systemicity and Scrutability 318

5. Concluding Remarks 320

9. SYSTEMATIZING 323
1. Classification and Theory 324
1.1. Classification 324
1.2. Theory 331

2. Construction and Reality 340
2.1. Theory Construction 340
2.2. Theory and Reality 350

3. Convention and Law 355
3.1. Convention 355
3.2. Law 359

4. Concluding Remarks 373

APPENDICES 377

1. The Power of Mathematics in Theory Construction: A Simple Model of Evolution 377

2. The Prose Identifying the Variables 380

BIBLIOGRAPHY 383

INDEX OF NAMES 396

INDEX OF SUBJECTS 401

PREFACE TO *EPISTEMOLOGY I & II*

We have learned a good deal about many things, we have come to realize that if we value knowledge we should continue to inquire because we shall never know enough, and we keep expanding, applying, and diffusing nearly every bit of knowledge we manage to acquire. So many people are engaged in generating new knowledge, using it, and diffusing it, that economists speak of the *knowledge industry.*

Yet we are still remarkably ignorant about knowledge itself. We still argue whether it is the brain or the mind that knows, and are uncertain about the cognitive abilities of subhuman animals as well as of computers. We have only the most rudimentary information about the development (ontogeny) and evolution (phylogeny) of such abilities, and we continue to ask whether the study of knowledge is a harmless armchair specialty or a useful cross-disciplinary field of inquiry, both factual and theoretical, scientific and philosophical. In short, while the knowledge industry is thriving, the science and philosophy of knowledge are still in bud.

There are several possible explanations for the underdevelopment of our knowledge of knowledge. One could be that epistemology is impossible: this is the view of the radical skeptics and a few others (e.g. Nelson, 1912). Another view is that epistemology and, in general, philosophy has nothing new left to say (e.g. Rorty, 1979).

I find such skepticism and pessimism unacceptable, in view of the rapidly expanding body of interesting epistemological problems raised by science, technology, and the humanities, as well as recent advances in the physiology, psychology, and sociology of perception, learning, and ideation. There must be an alternative explanation for the underdevelopment of epistemology. We may have to blame the conservatism of most philosophers, who stick to the same old problems and rely exclusively on ordinary knowledge as well as on the authority of classical *isms* instead of engaging in a scientific investigation of the problems of cognition and knowledge. If this is the case, then what epistemology needs is not more obituaries but a radical reorientation. I hope this work will contribute to such renewal. This book and its companion, namely Volume 6 of my *Treatise*, concern the basic traits of cognition and knowledge, as well as the basic principles guiding the acquisition, analysis and utilization of knowledge. Their joint title could well be *Principles of Inquiry*, or *An Inquiry into Inquiry.*

This volume is devoted to general epistemology and methodology; the next, to some epistemological and methodological problems arising in contemporary science and technology. Epistemology, or the theory of knowledge (Fr. *gnoséologie*, Ger. *Erkenntnistheorie*), is the field of research concerned with human knowledge in general—ordinary and scientific, intuitive and formal, pure and action-oriented. And methodology—not to be mistaken for methodics, or a set of methods or techniques—is the discipline that studies the principles of successful inquiry, whether in ordinary life, science, technology, or the humanities.

In this work epistemology is conceived as a merger of philosophy, psychology, and sociology: it describes and analyzes the various facets of human cognitive processes, whether successful or not, and whether or not they bear on everyday matters. Methodology too is descriptive and analytical, but in addition it is prescriptive or normative: it attempts to find out not only how people actually get to know but also how they ought to proceed in order to attain their cognitive goals. Thus both the epistemologist and the methodologist are supposed to describe and analyze experiment, but the methodologist is primarily interested in well designed experiment. In short, whereas epistemology is concerned with inquiry in general, the task of methodology is to find or perfect optimal inquiry strategies.

Yet, for all their differences, epistemology and methodology overlap considerably. Both deal with cognitive processes, so both deserve being called 'epistemology'. After all, it is the same people who succeed and fail in attempting to learn. Both studies should be pursued, and often at the same time, if we are to improve our understanding of the mechanisms of our successes and failures in our cognitive ventures. Fortunately, the two lines of inquiry into inquiry intertwine in the research of a new breed of cognitive psychologists. (See Nisbett and Ross, 1980.)

This book continues an old tradition or, rather, a whole fan of traditions started in ancient Greece and India. But at the same time this work departs from tradition with regard to method. It is hoped that our inquiry into inquiry will be closer to the cognitive sciences and, in general, closer to contemporary research, than to obsolete dogma. More particularly, we shall proceed as follows. We shall pick up the rich legacy of epistemological problems and hints (often optimistically called 'theories') bequeathed to us by the epistemological tradition. We shall enrich it with some of the problems and findings of contemporary scientific, technological and humanistic research, topping it with new hypotheses compatible with the science of the day—in particular neuroscience, psychology, and social science. And we shall elaborate and systematize the whole with the help of a few modest tools such as the concepts of set and function. However, in contradistinction with the former volumes in this *Treatise*, here we shall adopt a far more modest level of formalization. The result is a greater intelligibility—and length. (The formalizations have been put in parentheses and in the Appendices.)

Finally, we shall try and put our epistemological principles to the test: we shall check whether they account for the actual conduct of inquiry or whether they might help improve it. In other words, we want empirical confirmation for our epistemology, and we want it to be useful to research. And, far from confining our considerations to the physical sciences and technologies, we shall take all the sciences and technologies as our testing ground. This undertaking should prove far more rewarding, though also far harder, than writing obituaries of epistemology in the manner of Wittgenstein or Heidegger.

ACKNOWLEDGMENTS

I am grateful to the 25 generations of students in Argentina, Uruguay, the U.S.A., West Germany, Switzerland, México, and Canada, who took my courses in epistemology and methodology, and asked interesting questions or provided valuable information. I also thank the following persons for their advice, criticism, or encouragement: Joseph Agassi (Philosophy, Tel-Aviv University), Antonio M. Battro (Centro de Investigaciones Filosóficas, Buenos Aires), Hervé Barreau (Fondements des Sciences, Université Louis Pasteur, Strasbourg), Félix Córdoba Alva (Centro de Investigaciones Biológicas, La Paz, BCS, México), Mike Dillinger (Linguistics, McGill), Bernard Dubrovsky (Psychiatry, Physiology and Psychology, McGill), José Ferrater Mora (Philosophy, Bryn Mawr College), Manuel Garrido (Philosophy, Universidad Autónoma de Madrid), Alberto Hidalgo (Philosophy, Universidad de Oviedo), Rolfe Leary (Forest Experiment Station, St. Paul, Min.), James McGilvray (Philosophy, McGill), Claude Panaccio (Philosophy, Université du Québec à Trois Rivières), Michel Paradis (Linguistics, McGill), Rafael Pérez Pascual (Physics, UNAM, México), Miguel A. Quintanilla (Philosophy, Universidad de Salamanca), Osvaldo A. Reig (Estudios Ambientales, Universidad Simón Bolívar, Caracas), Marcel Roche (Estudio de la Ciencia, IVIC, Caracas), William R. Shea (Philosophy, McGill), Jane Stewart (Psychology, Concordia University, Montréal), José Félix Tobar (Philosophy, McGill) and Raimo Tuomela (Philosophy, University of Helsinki). I also appreciate the patience of Jan Hattink and Mrs N. Jones, both with the Reidel Publ. Co. I gratefully acknowledge a research grant from the Social Sciences and Humanities Research Council of Canada. Last, but not least, I thank my housekeeper Mrs Daniela Konieczny.

M. B.
June 1983

SPECIAL SYMBOLS

$\neg p$	*not-p*
$p \mathbin{\&} q$	*p and q* (*conjunction*)
$p \vee q$	*p or q* (*disjunction*)
$p \Rightarrow q$	If p, then q (*implication*)
$p \Leftrightarrow q$	If p, then q, and conversely (*iff*)
$A \vdash B$	Premises A *entail* conclusions B
$=_{df}$	Identical by *definition*
$\{x \mid Fx\}$	The *set* of objects possessing property F
$\emptyset$	The *empty set*
$a \in A$	Individual *a belongs to* set A
$A \subseteq B$	Set A is *included* in set B
$A \cup B$	The set of objects in A *or* in B
$A \cap B$	The set of objects in A *and* in B
$A - B$	The set of objects in A but *not in* B
$A \Delta B$	The set of objects in A *or* in B but *not in both*
$\langle a, b \rangle$	The *ordered pair* of a and b
$A \times B$	The *cartesian product of* A and B, i.e. the set of ordered pairs $\langle a, b \rangle$, where a is in A and b in B
2^A	The *power set* (family of all subsets) of set A
$\lvert A \rvert$	The *cardinality* (numerosity) of set A
$\mathbb{N}$	The set of *natural numbers* (0, 1, 2,...)
$\mathbb{R}$	The *real line*
$\mathbb{R}^+$	The set of *non-negative real numbers*
$f: A \rightarrow B$	The *function* f maps set A into set B
$f(x)$	The *value* of function f at x
$F = \langle F_1, F_2, \ldots, F_n, \ldots \rangle$	A state function for some system
R & D	Research and development
S & T	Science and technology

INTRODUCTION

In this Introduction we shall state the business of both descriptive and normative epistemology, and shall locate them in the map of learning. This must be done because epistemology has been pronounced dead, and methodology nonexisting; and because, when acknowledged at all, they are often misplaced.

1. DESCRIPTIVE EPISTEMOLOGY

The following problems are typical of classical epistemology:

(i) What can we know?
(ii) How do we know?
(iii) What, if anything, does the subject contribute to his knowledge?
(iv) What is truth?
(v) How can we recognize truth?
(vi) What is probable knowledge as opposed to certain knowledge?
(vii) Is there a priori knowledge, and if so of what?
(viii) How are knowledge and action related?
(ix) How are knowledge and language related?
(x) What is the status of concepts and propositions?

In some guise or other all of these problems are still with us. To be sure, if construed as a demand for an inventory of knowledge the first problem is not a philosophical one any more than the question 'What is there?'. But it is a genuine philosophical problem if construed thus: 'What kinds of object are knowable—and which ones are not?' However, it is doubtful that philosophy can offer a correct answer to this problem without the help of science and technology. For example, only these disciplines can tell us whether man can know not only phenomena (appearances) but also noumena (things in themselves or self-existing objects).

Problem (ii) used to be called that of the sources of knowledge: do we know by revelation, intuition, sense experience, action, or reasoning? It may now be reformulated as the problem of cognitive mechanisms, and it is of interest to psychology, not only to philosophy.

Problem (iii)—about the possible contribution of the subject—used to trigger automatic responses, in particular 'Nothing' (empiricism) and 'Everything' (rationalism). It can now be investigated with the help of psychology.

Problem (iv) is ill-conceived, for it presupposes that there is a single kind of truth. Actually there are at least four kinds of truth (or four kinds of proposition to which a truth value can be assigned): logical, mathematical, factual, and philosophical. Logical and mathematical truth have become the property of logic and metamathematics. The problem of philosophical truth belongs to both descriptive and normative epistemology. And truth of fact, the concept occurring in all inquiries into matters of fact, is investigated by semantics—and perhaps can also be approached by psychobiology.

Problem (v) is that of truth tests and criteria, and is therefore best left to normative epistemology or methodology.

Problem (vi) is that of the fallibility of factual knowledge—or, to put it in positive terms, of the corrigibility of hypotheses and data. And it may or may not be linked to the probability calculus. (It is not so linked in the present work.)

Problem (vii) is that of the scope of pure reason, the nature of formal science, and the existence of (true) a priori statements of fact—hence it is actually a whole family of problems.

Problem (viii) is central to the philosophy of technology as well as to dialectical materialism and pragmatism. Any examination of it should include an analysis of the role of knowledge in the design and implementation of plans, as well as a discussion of the value of praxis in posing cognitive problems or even in evaluating truth claims.

Problem (ix) is at the very core of the philosophies of language and mind. However, it should be handled with an eye on developmental psychology, neurolinguistics, and psycholinguistics.

Finally problem (x) belongs also to semantics and ontology, as well as to the philosophy of logic and mathematics.

Note that all of the above problems continue to draw the attention of some people and may thus be said (metaphorically) to be alive. (A problem is "dead" if no living being deals with it.) And they are "alive" because they continue to be open, i.e. not fully solved. Note also that epistemology has ceased to hold a monopoly on those problems: it shares them with other disciplines, be they other branches of philosophy or special sciences. In other words epistemology is (if not *de facto* at least *de jure*) no longer an independent discipline. The moral is clear: Do not pretend that epistemological problems can be attacked without the help of allied disciplines, in particular (though not exclusively) semantics, psychology and social science.

Nor are the above the only epistemological problems left, not even the only classical ones. Here is a short and haphazard list of problems on which epistemologists and sociologists of knowledge are currently working:

(xi) What is the role of mathematics in the generation of factual knowledge?
(xii) What is rationality?
(xiii) Can belief and plausible reasoning be regimented or formalized?
(xiv) What are problems, methods, approaches, hypotheses, theories, and rules?
(xv) What are the peculiarities of tacit knowledge (*know how*) and of explicit knowledge (*know that*)?
(xvi) What are the commonalities and peculiarities of description, classification, explanation, and prediction?
(xvii) What are basic science, applied science, and technology?
(xviii) What, if any, is the role of morality in inquiry?
(xix) How does the social matrix influence the cognitive activities?
(xx) What are the characteristics of learning (or inquiring) communities?

Note that, whereas traditional epistemology focused on knowledge, modern epistemology is also interested in the applications of knowledge. Second, whereas classical epistemology was concerned solely with the knowing subject and his (never her) accomplishments, or both, modern epistemology is also interested in learning communities as well as in the social function of cognition and in the social constraints on it. Consequently some of its problems are shared by social science, in particular the sociology of knowledge and social psychology. Recalling our comments on some of the typical problems of traditional epistemology, we must conclude that epistemology is only one of the cognitive sciences and it has a considerable overlap with its sister disciplines, in particular psychology (cognitive, developmental and evolutionary) and sociology (in particular social psychology and the sociology of knowledge).

2. NORMATIVE EPISTEMOLOGY (METHODOLOGY)

Unlike descriptive epistemology its normative partner, methodology, has only weak roots in Antiquity and the Middle Ages. Leaving logic aside, they amount to a handful of rules, such as Plato's injunction to shun opinion (*doxa*) and seek only certain knowledge (*episteme*), Aristotle's practice of defining everything, Hippocrates' recommendation to abstain from supernaturalistic explanation, and Ptolemy's advice that astronomy be restricted to describing phenomena (appearances).

The idea of a universal method for acquiring knowledge did not appear until *ca.* 1600. Discoursing on method became then fashionable, even a craze. There was hardly a scholar who doubted that, if only the right method could be found, mankind would attain instant wisdom. Bacon swore by induction and Descartes by conceptual analysis and deduction from indubitable first principles. Leibniz dreamed not only of a logical algorithm (*characteristica universalis*) but also of an *ars inveniendi* leading securely to first principles. It was a time of mythodology and methodolatry rather than methodology. The illusion lasted nearly three centuries.

Methodology proper did not come until the first modern philosophers of science, in particular Whewell (1847). Although some investigators deny the existence of methodology, it is easy to find methodological problems, many of which have not yet been given satisfactory solutions, or at least generally accepted ones. Here is a sample of rather classical problems of this kind.

(i) Is there a single best way of advancing knowledge and, if so, which is it?
(ii) What is the scientific method?
(iii) What kinds of convention do we stipulate and which are their roles in inquiry?
(iv) What are the rules of correct definition? And are there creative definitions?
(v) What are the principles of correct classification?
(vi) What is the role of regulative (or heuristic) principles in inquiry?
(vii) What kinds of theory should we prefer: descriptive (phenomenological) or explanatory (mechanismic), deterministic or probabilistic, single-level or multilevel?
(viii) What is the point of axiomatics?
(ix) What is testability: confirmability, refutability, or either?

(x) If theories are tested as wholes, how can we spot their false components?

All of these problems are still topical and some of them are still open. Whereas some of them, in particular the first two, are discussed even in bull sessions, others are the subject of technical papers or books. But all of them are likely to be faced, sooner or later, by all investigators. For example (v), or Aristotle's problem, is very much alive among biologists, particularly in connection with numerical and evolutionary taxonomy. Problem (vi), Kant's, comes up in every discussion of correspondence principles. Problem (vii) is confronted by anyone who proposes to build a factual theory or model. Problem (viii), Hilbert's, is discussed by both friends and foes of axiomatics. Problem (ix) is the bone of contention between confirmationists and refutationists. And Duhem's problem (x) is faced by any investigator in the presence of some evidence against an otherwise well confirmed theory.

Here is a random sample of methodological problems that have come more recently to the fore:

(xi) Are properties dispensable? If not, how are they to be represented?
(xii) Can there be any techniques of theory construction?
(xiii) What are the advantages of the reduction (deduction) and of the merger (synthesis) of theories and entire disciplines?
(xiv) How should hypotheses and theories be evaluated? In particular, is empirical confirmation necessary and sufficient?
(xv) Do we have to test all the alternative hypotheses that might account for a given group of facts, or only those enjoying some empirical or theoretical support?
(xvi) How are revolutionary new ideas or procedures to be evaluated? Do they have to be compatible with some of our previous knowledge or can they be totally deviant?
(xvii) What would be the value of measurement and experiment if the experimenter had powers of psychokinesis and precognition?
(xviii) What can be the value of empirical tests if they are designed with the help of theories?
(xix) How should methods be tested?
(xx) How should the rules of statistical testing (e.g. for significance or for randomness) be tested?

Problem (xi), prompted by the nominalist mistrust of properties and their conceptualization (attributes), is both philosophically and technically interesting. Problem (xii) was initially raised by the believers in the possibility of fashioning recipes for inventing, and has recently been revived by some computer enthusiasts. (If empiricists are right in holding that theories are data packages or distillates, why not build or program an electronic theorist?) Problem (xiii) comprises the questions of reductionism and of the unity of science. Problem (xiv) questions the empiricist tenet that empirical confirmation is not only necessary but also sufficient to canonize hypotheses and theories. Problem (xv) questions the refutationist conviction that experimental scientists have all the time in the world to try and refute any hypotheses or theories that may occur to us. Problem (xvi) is that of weighing scientific revolutions and distinguishing them from crackpot proposals. Problem (xvii) suggests that experimental science presupposes the falsity of parapsychology—which in turn poses the problem of the empirical testability of belief in the paranormal. Problem (xviii) amounts to this: Are empirical procedures contaminated by theory to the point that they have no test power? And the last two problems pose the difficult question of evaluating the goodness of methods and rules.

Let the above sample suffice to show that methodology can be neglected only at the risk of making bad mistakes. And note the salient characteristic of methodological problems in contrast to problems in descriptive epistemology (Section 1). Whereas the latter are questions of fact—How do we know X?, What is cognitive item Y?, and the like—methodology handles typically value problems—What is X worth?, What is the correct way of doing Y?, and the like. Consequently, whereas an epistemological problem is answered by a description, a definition, or a theory, a methodological problem is answered, in the last analysis, by a value judgment or by a rule of procedure—either of which must in turn be justified in terms of theory, experiment, or praxis. In short, methodology, like logic and decision theory, is a normative discipline.

3. EPISTEMOLOGY AND BIOLOGY

All cognitive activities, from sniffing and exploring to theorizing and forecasting, are biological functions—as well as social ones in the case of social animals. They are all aspects of the adaptation of animals to their environment as well as of their activity in altering their environments to suit their needs. Hence biology is, or ought to be, interested in cognition, and all the cognitive sciences are, or ought to be, based on biology.

The view that cognition is a biological process emerged in the late 19th century, which saw the beginnings of experimental psychology as well as of comparative (or evolutionary) psychology, later to be supplemented by developmental (or genetic) psychology. Perception ceased to be a subject of philosophical speculation to become a problem for physiologists and experimental psychologists. Even problem solving and other higher mental functions became eventually the object of experimental investigation and, more timidly, mathematical modelling. Darwin, Helmholtz, Mach, Lloyd Morgan, Yerkes, Köhler, Piaget, Békesy and Hebb were among the pioneers of the biological approach to psychology and, in particular, the study of cognitive abilities. Today this study is pursued vigorously by physiological psychology (cf. Hebb, 1966; Milner, 1970; Bindra, 1976). This approach to cognition must not be confused with *cognitivism*, or the information processing view, just as the study of behavior must not be identified with behaviorism. Indeed the gist of cognitivism is that biology is irrelevant to the understanding of cognition, whereas computer science is essential to it (cf. Simon, 1979; Pylyshyn, 1980).

The *rapprochement* between psychology and epistemology, on the one hand, and biology on the other, was not limited to the scientific community but fascinated also some philosophers. Thus Nietzsche, the real founder of pragmatism, held that lies may be just as useful for life as truths, and claimed that the sole justification for research lies in its usefulness to preserve life. Likewise Spencer is the author of the famous dictum "Science is for life, not life for science". (Many contemporary statesmen and bureaucrats hold the same narrow-minded view of science without having read either Nietszche or Spencer: we keep reinventing falsity.) Spencer also wrote the celebrated sentence echoed in our days by Konrad Lorenz: "What is *a priori* for an individual is *a posteriori* for a species". (That is, the individual does not have to start from scratch, for it inherits a number of traits won in the course of evolution.) And Mach—noted for his important discovery that sensation is a function not only of stimulation but also of inhibition—wrote occasionally of the "transformations and adaptations" of thought. He also wrote about "the adaptation of thought to sensation" (which he regarded as the sole firm source of knowledge), as well as about the conflict between the two (for him the sole source of problems). But he did not spell out how cognitive behavior can be beneficial or harmful, let alone how selection could uproot incorrect cognitive habits.

Neither Nietzsche nor Spencer nor Mach may be said to have built a biological psychology, let alone an evolutionary one. (See Čapek (1968) for the opposite view.) The same holds for Peirce, another thinker influenced by Darwin's theory, and who stated that the truest beliefs are those that "live" longest—an idea refuted by the persistence of superstition. Nor is evolutionary epistemology related to Popper's analogy between natural selection and refutation, let alone to his speculations on disembodied "world 3" items (Popper, 1972). Indeed the theory of evolution applies only to concrete systems, namely biopopulations, not to disembodied objects such as ideas in themselves (Bunge, 1981a).

If epistemology is to become scientific it will have to learn not only from neurophysiology and physiological psychology but also from developmental (genetic) psychology and from whatever evolutionary biology has to say about cognition. That is, it will have to include evolutionary epistemology. So far this is little else than a thesis and a program. The thesis is this: "Our cognitive apparatus is a result of evolution. The subjective cognitive structures are adapted to the world because they have evolved, in the course of evolution, in adaptation to that world. And they match (partially) the real structures because only such matching has made such survival possible" (Vollmer, 1975, p. 102). This is a programmatic thesis: it orients a whole research project that has still to be implemented not only with more observations in comparative psychology and ethology but also with theoretical synthesis and speculation, as well as the conceptual clarification (of, e.g. the hazy notions of cognitive structure and real structure). However immature, the project of evolutionary epistemology is well grounded, not wild: it fits in with evolutionary biology.

4. EPISTEMOLOGY AND PSYCHOLOGY

Except for physiological psychologists and the rare zoologist, ethologist, or neuroscientist, cognition is usually regarded as a suprabiological phenomenon and studied separately from the nervous system. One main reason for this non-biological approach is the persistence of the ancient myth of the immaterial mind. This myth is not only a component of most religions and popular world views but also of most psychological doctrines (in particular psychoanalysis) and most philosophies of mind. (For a defense see Popper and Eccles (1977), and Eccles (1980); for criticisms see Bindra (1980), Bunge (1980a, 1981a), and Hebb (1980).)

The detachment of cognition from its organ, namely the nervous system, deprives psychology and epistemology of substance and depth: it is like discoursing about the weather as if it were something different from the state of the atmosphere. By the same token it prevents psychology and epistemology from benefiting from the findings of neuroscience and, in general, it isolates them from all the other cognitive fields. (Isolation is characteristic of protoscience and pseudoscience.)

If mind and matter are distinct substances, then it is hard to understand how the former can get to know the latter, except by adopting some form of idealism. In this case matter does not exist except as a figment of our own mind, and the acquiring of knowledge about anything is reduced to the mind knowing itself or other minds, whether individual or supraindividual. This may work for logic and mathematics but it fails to account for natural science and technology, as in these fields one must take the real world for granted if one wishes to learn something about it. In fact, there would be no point in exploring the world unless it were out there all by itself. In particular we would waste no time experimenting.

On the other hand materialism has no trouble understanding, at least in principle, how a material system (the brain) can get to know something about other material things or even about itself. Indeed, according to materialism (and physiological psychology) every cognitive process is a string of events in some central nervous system. So it may be imagined that it can match other material processes, much as an electrocardiograph transduces voltage variations into graphs, or a radio receiver transforms electromagnetic waves into sound waves. Moreover, materialism happens to be the ontology of modern science, which, far from countenancing disembodied or ghostly entities, assumes that all existents are lawfully changing concrete (or material) things, though not necessarily of the same kind. (See *Treatise*, Vols. 3 and 4.) For these reasons we choose materialism, and more particularly scientific materialism (Bunge, 1981a), as the general ontological view which our epistemology is to fit.

This approach to the study of cognition contrasts with both traditional mentalism and neomentalism, or the philosophy of mind adopted by most workers in the so-called *cognitive science*. This is the collective name of three disciplines concerned with cognition: cognitive psychology, linguistics, and artificial intelligence. Whatever the accomplishments of these fields, they share an important shortcoming, namely, that all three ignore the nervous system and are therefore isolated from neuroscience and psychobiology. Furthermore, cognitivism has produced no new theoretical models and it has ignored motivation and response production, so it cannot possibly offer a comprehensive framework for psychology (Bindra, 1982). In particular, talk about human information processing and computing is just a metaphor and, as such, no substitute for a theory. (Recall Hobbes' characterization of metaphor as *ignis fatuus.*)

Nor is our approach in sympathy with behavioristic psychology. The investigation of cognitive processes includes the observation and measurement of behavior but cannot be restricted to it because the motivations, intentions, cogitations, goals and plans of the inquiring subjects are not observable. Besides, what the inquirers do is totally unintelligible to an observer who does not have a similar background. For example, a chemist, a chemical engineer and an alchemist may wear white laboratory coats, handle roughly the same instruments and the same chemicals, and even go through similar motions, while doing entirely different things because they think differently and have different goals and methods. For this reason the fashionable anthropological and sociological studies of the scientific community (e.g. Latour and Woolgar, 1979), when limited to observing the behavior of scientists, are useless when not positively damaging to the science of science.

5. THE PLACE OF EPISTEMOLOGY

Many disciplines besides descriptive and normative epistemology study cognition and its outcome, i.e. knowledge. Zoology and ethology are interested in cognition; so are neuroscience and psychobiology (or biopsychology), as well as the non-biological approaches to cognition, in particular neomentalism, linguistics, and artificial intelligence. Nor are these the only members of the family that could bear the hybrid name *cognitology.* All of the social sciences are interested in cognition and knowledge: in the former because learning is largely a social process, in the latter because knowledge is a social force—or, rather, people who know something out of the ordinary can affect the behavior of others. In particular, anthropologists study the functions of the learning professionals, be they shamans or molecular biologists; archaeologists conjecture what ancient peoples must have known in order to manufacture the artifacts that are dug out; social psychologists investigate the social constraints on cognition as well as the effects of knowledge on social structure; and historians—particularly historians of science, technology, or the humanities—are interested in the traditions and revolutions in modes of inquiry and belief, as well as their relations with environmental, demographic, economic and political factors.

All the members of cognitology have a common referent, namely the inquiring system (individual or group). Therefore there is—inevitably and happily—some overlap between the various sciences of cognition and knowledge. However, a common reference does not ensure identity, for one and the same subject matter can be studied from different viewpoints—i.e. one can ask questions of different kinds about one and the same entity. This holds, in particular, for the scientific and the philosophical approaches to cognition and knowledge. For example, whereas the scientist may study perception and perceptual illusion as sources of knowledge and error respectively, the philosopher may also study the scope of perception, the nature of perceptual knowledge and its differences from conceptual knowledge, as well as the contrast between appearances (as presented in perception) and reality (as conjectured by theory). Whereas the scientist may be interested in the way children come to know about objective constancies (e.g. conservation laws), the philosopher may puzzle the nature of law statements, their relation to objective patterns, and the role of such statements in science and technology. And whereas the scientist may investigate the origin—psychological or historical—of theories, the philosopher may study the very nature and role of finished theories, as well as the conceptual (rather than psychological or cultural) differences between them.

Both approaches to cognition and knowledge, the scientific and the philosophical, are necessary. The scientist could not even start studying, say, the process of hypothesis formation (in some brain and under certain social circumstances) unless he had a reasonably precise and correct idea of what a hypothesis is—which idea it behoves philosophers to exactify. And the philosopher should be happy to learn that somebody studies the genesis, evolution, and demise of hypotheses, not only in highly educated human adults but also in children, primitive man, and even in some subhuman species.

Cognition and knowledge can then, nay must, be studied *both* scientifically and philosophically. The two approaches are needed because they are mutually complementary and they stimulate and control one another. (Only bad science is the enemy of good philosophy, and bad philosophy the rival of good science.) Moreover there is no clear demarcation between scientific and philosophical epistemology, and none should be invented.

(Likewise there is no borderline between science and ontology: see *Treatise*, Vol. 3, Introduction.) This way of conceiving of epistemology is likely to advance our knowledge of cognition and knowledge, for each partial view can be allowed to enrich and check the other.

An additional advantage of conceiving of philosophical epistemology as a member of the large family of cognitive sciences (or *cognitology*) is this. Ordinarily epistemology is regarded as being far removed from both ontology and ethics. In our perspective cognition is only a special kind of biological process, and therefore it is an object of study of the ontology of organisms; and organisms capable of sharing knowledge are members of some society or other, and therefore the object of study of the ontology of society. Thus any reasonably true generalizations about cognition may be placed in the intersection of the science and the philosophy of cognition.

Nor is epistemology alien to ethics or the study of morality. Indeed normative epistemology is about rules of correct (or successful) inquiry; and rules, unlike laws of nature, can be either obeyed or broken with honest or dishonest intentions. Besides, all knowledge is valuable in one way or other; and anything valuable can be well used or ill used, shared or stolen—as well as traded and faked. The pursuit of knowledge must then involve a moral code. In fact it does involve one: recall the commandments to prefer truth to error, to check conjectures, not to make up data, to seek criticism, to acknowledge truth in the rival and error in the friend, to share information, and so on. This code of inquiry intersects with methodology: every correct procedure of inquiry is intellectually honest even though it may be morally objectionable on other grounds.

What about logic and mathematics in their relation to epistemology? These two formal sciences provide the backbone of all the other sciences but they do not investigate cognition and they are not in a position to discriminate factual truth from factual error. In short, they are not members of cognitology. To be sure, logic used to be regarded as the *organon* of inquiry and thus as identical with methodology. But we have learned since the beginning of the modern period that logic may at most discover (logical) error: it is not a guide of inquiry but a constraint on it. We now tend to conceive of logic as the intersection of mathematics and philosophy, as well as a tool of conceptual analysis and criticism in all cognitive fields.

In conclusion, epistemology is not an independent discipline. It must make use of logic and can make use of mathematics, and it belongs to the family of cognitive sciences. Moreover it has considerable overlaps with all

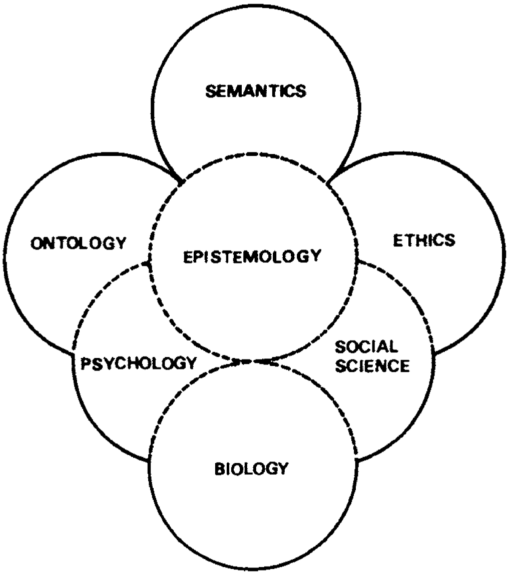


Fig. 0.1. The place of epistemology among the cognitive and philosophical sciences. Note the overlappings with neighbouring fields and the lack of clear cut boundaries between them.

of these disciplines as well as with ontology and ethics, not to mention semantics, from which it is often indistinguishable. See Figure 0.1.

6. USES OF EPISTEMOLOGY

There should be no need to justify research on cognition and knowledge: only barbarians or yahoos would question the worth of inquiring for the sake of knowing. But since scientists and philosophers are besieged by barbarians and yahoos, it won't hurt to argue for the practical uses of such studies.

To begin with, learning theory, though still in its infancy, has already had a profound impact on the theory and practice of teaching. (See Suppes (Ed.), 1978.) Secondly, behavior therapists are having some success in correcting bad habits, in particular bad cognitive habits, on the basis of the hypothesis that whatever has been learned can be unlearned. (See Bandura, 1974; Wolpe, 1976, 1978.) Thirdly, neurologists have started to cope with learning disabilities attributable to congenital neural defects or brain injuries. (See Hécaen and Albert, 1978; Pincus and Tucker, 1978.) In short the psychology of cognition, though still very young, has proved its practical usefulness. Much more can be expected from it if it sheds the myth of the immaterial mind as well as the superficiality and dogmatism of behaviorism.

The social studies of cognition and knowledge have made important inroads as well (cf. Merton, 1973). The most important of them has been showing (not just asserting) that social circumstances condition our mode of perceiving, thinking, and evaluating, so that cognition and knowledge cannot be adequately understood except in their social context. Moreover it has now become clear that, at least in man, knowledge and sociality are coextensive: neither exists without the other. This discovery has had a profound impact on the history of science and technology, which until relatively recently followed the pattern of the history of ideas, investigating ideas in a social vacuum. (Inevitably, the discovery of the social dimension of cognition has led to sociologism, or the neglect of the brain as well as individual circumstances.)

Philosophical epistemology is a little harder to defend nowadays. However, if we care for science, technology, and the humanities, we should also care for their epistemological presuppositions—such as that there is an independent reality and that it can be known if only in part. We should dig up, cleanse, analyze, and systematize such principles. (For a short list see Ch. 15, Section 3.1.) It would seem obvious that, the better we know how we can get to know, the better we can improve (or block) the learning process, particularly in science, technology, and the humanities. So, a study of the epistemological presuppositions of research, as well as of any other tacit assumptions of scientific and technological research, should pay off in practical results.

As for normative epistemology, it is often claimed that it is no more needed for doing science or technology than a knowledge of grammar is needed for writing well or of logic for thinking correctly. True, some people untutored in logic and grammar think and write well, and most successful scientists and technologists have not been exposed to methodology. However, it is also true that the study of grammar improves good writing, that of logic good thinking, and that of methodology good research. Most people untutored in logic and methodology are prone to incur fallacies and find it difficult to design research strategies even for solving simple problems. In his classic *Thinking* (1958) Bartlett concluded that "especially with naive, although intelligent searchers, it is most likely that the method of search adopted will be an uneconomical one".

However, most scientists and technologists are indifferent to, when not contemptuous of, methodology. Philosophers tend to impute this attitude to excessive specialization, lack of philosophical curiosity, or even illiteracy. Scientists and technologists feel on the other hand that methodologists misrepresent what scientists and technologists do or ought to do (e.g. Medawar, 1969). Actually the situation is much worse than either party thinks, for *both* are right. Indeed there seem to be few if any persons capable of giving a truthful and balanced account of the actual conduct of inquiry—even of their own. Author *A* believes only in logic, *B* only in intuition; *C* swears by induction and *D* by deduction; *E* by experiment and *F* by theory; author *G* is a gradualist and *H* a catastrophist with regard to the evolution of knowledge; thinker *J* ignores society altogether and *K* places knowledge in society—and so on and so forth. Every one of them can cite examples in his favor, so neither of them is totally wrong. But none of them has built a system accommodating the many and apparently conflicting features of inquiry: each of them offers a one-sided view, and each is anxious to have it accepted as his own original *ism*.

We must try and strike a golden mean between *methodolatry* and *methodoclasm*. The former is the view that method is more important than product. (Paradigm: the Chinese dictum "Do not give me a fish: instead, teach me how to fish.") Extreme methodolatry will promote the teaching of methods separate from subject matter, and so the student will learn neither. On the other hand methodoclasm holds that only knowledge is important, not how we come to know. (Paradigm: "What matters with a piece of news is whether it is true or false, not the means—hearsay, reading, intuition, etc.—we come by it.") Extreme methodoclasm will promote teaching the results of research with neglect of the way research was initiated, planned, conducted, and evaluated.

A balanced view will hold that both the process of inquiry and its product are worthy of study. This view has two advantages. One is that it allows one to examine and evaluate methods in the light of results and conversely—the virtuous circle principle. (See Bunge, 1962; Rescher, 1977.) Another advantage is that a study of methodology serves to rectify a number of misconceptions about inquiry. Some scientists have been known to be busy on problems other than those they believed they were investigating; at a loss to formulate the most obvious hypotheses or even reluctant to make any conjectures; incapable of clarifying key notions they used daily; totally uninterested in neighboring fields; ignorant of the ultimate goals of scientific research; contemptuous of theory or, more rarely, of empirical test; believing that all the mathematics they need is a handful of statistical recipes—or, on the contrary, that mathematics suffices for science. Any such scientist would profit by some methodology. And even the most experienced and successful researchers can benefit from pondering, once in a while, over such rules as: "Place your problem in a wider context", "Look for deeper (non-apparent) variables", and "Do not underrate chance". (See Ch. 15, Section 3.2, for a short list of regulative principles of research.)

7. CONCLUDING REMARKS

Epistemology teems with open problems: a sign that it has a right to life. But epistemology is still immature: a sign that it has a long way to go. It cannot go it alone and in fact it has interacted vigorously with natural science, in particular physics, since the 17th century. But until recently epistemology has had only very weak ties with evolutionary biology, neuropsychology, and social science—not surprisingly, for these came long after epistemology was born.

Epistemology must mesh in with biology, psychology, and social science, because the learning process happens to occur in and among animals. The immaterial and isolated (and male and adult) knowing subject of traditional epistemology must be replaced with the inquiring brain, or team of brains, embedded in society. The unchanging mature mind must be replaced with developing and evolving animals exploring their environment as well as themselves, and changing both with the help of the knowledge gained in such inquiries. Epistemology, in sum, must be biologized and sociologized.

However, this proposed reconstruction of epistemology with the help of biology, psychology, and social science, should not be taken as a pretext for neglecting the epistemological problems that can be handled independently. Thus regardless of whether or not conjecturing and jumping to conclusions are inborn human dispositions; regardless of the particular neural systems that do the conjecturing and the jumping to conclusions; and regardless of the way social circumstances stimulate or inhibit such brain processes, hypotheses and inferences have certain logical, semantical, and methodological properties (e.g. generality, factual reference, and testability) that can and must be studied separately from the corresponding concrete biological and social processes. Likewise we do not need biology, psychology or social science to check whether a computation is correct or whether the components of an electric circuit have been assembled correctly.

In short, up to a certain point methodological inquiry can be pursued independently of descriptive epistemology. This relative independence explains why we have been able to learn (and stipulate) something about the *correct* way of going about forming and combining concepts and propositions, classing, comparing, analyzing, etc. without really knowing much about the neurophysiology (or even the global psychology) of such operations. Nevertheless this situation is anomalous and reminds one of the engineering feats of the ancients, which had no scientific basis, rather than of modern technology, which is science-based. We may conjecture that a more intense and deeper scientific investigation of cognition, and in particular of the neurophysiology of thinking, may bring about a revolution in methodology or at least a more intimate understanding of the relation between the *is* and the *ought* of knowledge. But for the time being we are faced with the relative independence and more advanced state of development of normative epistemology. This explains why it gets the lion's share in this book.

PART I

COGNITION AND COMMUNICATION

CHAPTER 1

COGNITION

This book is about knowing and knowledge, error and ignorance: about the cognitive process, its pitfalls, and its outcome. When one perceives or conceives an object, investigates it, manipulates it or evaluates it, one engages in a cognitive process. With some ability and luck—i.e. under suitable circumstances—one will learn something about the object of inquiry. Whatever one has learned in this process, added to what he may have known beforehand, is one's *knowledge* of the object.

The distinction between a cognitive process—in particular an inquiry—and its result at a given moment is an instance of the process-state distinction. (A process may be construed as a sequence of states: Vol. 3, Ch. 5, Section 2.3.1.) It is a useful distinction because, for certain purposes, the end result is of little interest, whereas for others it is all that matters. Thus the neuroscientist and the psychologist are usually more interested in the learning process—and its disruptions—than in its accomplishments, whereas the latter attract most of the attention of traditional philosophers as well as of historians. Here we shall be concerned with both.

We ask indeed *Who* (or *what*) *can know what, and how*? This complex question summarizes the problematics of all the disciplines—philosophical, scientific and technological—concerned with cognition or with knowledge. The first part of the question, namely *who* (or *what*) knows, is handled by psychology (both animal and human) as well as by the sociology and history of knowledge. The second part, i.e. *what* can get known, interests epistemology and the history of science, technology, and the humanities. And the third part, namely *how* (by what means, mechanisms, or methods) we may get to know, is addressed by all the cognitive sciences, particularly methodology.

There are three fruitful approaches to the study of cognition: the neurobiological, the psychological, and the philosophical ones. All three are concerned, each in its own way, with the same problematics: internal representation, imagery, conception, conjecturing, systematization, inference, problem solving, criticism, and their kin. The neurobiological study of cognition attempts to lay bare the neural mechanisms of cognitive processes. The psychological study tries to describe them at the molar or phenomenological level, i.e. without inquiring into their neurophysiology. And the philosophical (epistemological) study endeavors to elucidate those concepts, as well as to discover the contribution of the various modes of knowing to the common enterprise, which is the perfectible knowledge of the world, including ourselves and our cognitive abilities.

Of these three approaches the oldest is the philosophical one, and it is fair to say that philosophers have so far contributed no less to the science of cognition than either psychologists or neuroscientists. But it is also fair to say that the philosophy of cognition is now at the end of its tether: one can only grasp so much of reality without the help of experiment and modeling. Psychology took up the problem of cognition ever since it became an independent science. True, neither behaviorists nor psychoanalysts placed much value on investigating cognition, and both schools ignored that which does the knowing, namely the central nervous system. But other schools and fields, in particular primatology, comparative psychology, developmental psychology, and the Gestalt group, continued to work on cognition.

Interest in cognition has been stronger than ever since about 1960, both among psychologists and neuroscientists. However, many of the former have adopted a new form of mentalism, where analogies with artificial information processors are central whereas the nervous system is all but ignored. They hold cognition to be a matter of software (mind) not hardware (brain). We shall ignore this new version of the old doctrine of the soul and shall adopt instead the psychobiological approach sketched in Vol. 4, Ch. 4. (See also Bindra, 1976; Bunge, 1980a; Hebb, 1980.)

According to psychobiology (or neuropsychology) every mental function is a brain function, but the converse is false. (There are "household" functions, such as metabolism and the synthesis of proteins and hormones, which are nonmental.) This view unites psychology with biology and suggests a vast research project, both experimental and theoretical, that is only beginning to explain the mind and, in particular, cognition. (The project is not new, for it is found already in Henry Maudsley's *The Physiology of Mind* (1876), which gained a considerable readership but few followers.) Because this project is only in its initial stage, many of its hypotheses and theories are speculative, and it offers many more problems than solutions—a sure sign of youth in science. So, it should be judged not so much for its accomplishments as by its promise and philosophical soundness, as well as by the failure of its alternatives, namely mentalism and behaviorism.

1. The knowing subject

1.1 *The Cognitive Organ*

Looking for external objects and perceiving some of them (or forming the idea that we have failed to do so); watching for internal bodily signs and feeling them (or realizing that we failed to perceive them); learning motor skills and intellectual ones; forming concepts or propositions, questions or commands; identifying and classing, inferring and computing, criticizing and theorizing; puzzling and pondering, expecting and planning; speaking and listening to talk—all these and many more are *cognitive* acts or processes. They all generate, transform or utilize bits of knowledge.

Our initial assumption is that *every cognitive act is a process in some nervous system*, whether human or not. To put it negatively: *There is no knowledge in itself*, i.e. separate from the cognitive processes occurring in some nervous system or other. In particular there are no ideas except in brains of certain kinds—presumably those of mammals and birds. No autonomous ideas, i.e. ideas detached from brains, have ever been found and, if psychobiology is right, there can be none.

Our initial assumption is of course at odds with the idealistic tradition. We adopt the naturalistic (materialist) view that Platonic ideas, ideas in themselves (Bolzano, 1837), objective knowledge without a knowing subject (Popper, 1972), and objective mathematical reality (Putnam, 1975) are so many figments of the metaphysician's brain (Bunge, 1981a). Just as there is no motion apart from moving things, so there are no ideas in themselves but, instead, ideating brains. To be sure we may *feign* that there are ideas in themselves and in fact we must often make such pretense. We do so whenever we abstract from the real people who think up such ideas as well as from the personal and social circumstances under which they ideate. We may do so as long as we do not forget that we indulge in fiction, whether useful or idle. If in doubt that this is indeed a fiction, imagine the fate of human ideas upon a total nuclear holocaust.

All mental processes are cognitive operations but the converse is false. Thus reading and writing, drawing and often also calculating, involve sensory-motor activities as well as the manipulation of artifacts, in particular learning tools. Over the past four hundred years man has invented and manufactured thousands of learning tools, from microscopes to thermometers, from particle accelerators to high speed computers. All these artifacts are used as adjuncts to the nervous system of somebody, expert or lay: they enjoy no autonomy. Even an automated measuring device presupposes a design involving hypothesizing, and its supplies food for further thought; and even the most automatic of computations is but a link in a cognitive process occurring in some brain. Just as there is no robot work but only robot-aided work, so there is no automatic cognition but only computer-aided cognition.

Since cognition is a function of the nervous system, we must turn to neuroscience and psychobiology (physiological, developmental and evolutionary psychology) for its explanation. (Pure psychology, traditional or modern, can only supply some data and pose some problems about cognition. Genuine scientific explanation, unlike mere description and prediction, involves some mechanism or other—in this case some neural mechanism.) Now, the psychobiological study of learning suggests that learning consists in a change in the connectivity of some neuronal system or other. To put it negatively: If a behavior is unchanging (stereotyped, genetically programmed), then it is controlled by some system composed of neurons the interconnections of which do not change quickly over time. (To be sure these rigid systems are not unchangeable: they develop from birth on and deteriorate with old age. Moreover, if injured or even destroyed they can often reassemble somewhere else. But the point is that these changes are slow compared to the extremely fast changes in neuronal connectivity that we identify with mental processes.) In short, only *plastic* (or modifiable, or self-organizable, or uncommitted) neuronal systems are capable of learning anything. We call any such system a *psychon* (Vol. 4, Ch. 4).

Our first postulate, then, is that learning is an activity of plastic neuronal systems, or psychons. Our second assumption is that some animals do have psychons. (Exactly which species have them is for comparative psychobiology, or evolutionary psychology, to find out.) Our third, that all the psychons of an animal are coupled to one another forming a supersystem. Our fourth, that every animal endowed with psychons is capable of acquiring new biofunctions in the course of its life. (Apparently in man the processes of dendritic sprouting and formation of new synaptic connections stop only at the onset of senility. So, barring senility—which only about one out of five of us experiences—we can continue to learn new items throughout life.) Finally we decide to call *learned* any neural function involving a psychon that has acquired a regular connectivity, i.e. one that is constant or varies regularly rather than at random. A new connectivity may be formed by chance for the first time. If it consolidates, i.e. if there is recurrence or recall, we regard it as established or learned.

Imagine a group of neurons in a young child's cortex. There are three possibilities: (*a*) the neurons are committed, i.e. they form a (rigid) system or part of one, that is genetically programmed; (*b*) the neurons are available: they can associate in one or more ways, ephemerally or permanently, either in a constant manner or in a regularly varying one; (*c*) such association has already occurred. In the first case we say that the group of neurons in question has a rigid connectivity and therefore *cannot learn* anything; in the second, that it has a potential plastic connectivity and therefore *can learn* something; in the third, that it has already *learned* something, though it can presumably learn further items or unlearn some or all of them.

New neuronal systems can be formed during the lifetime of any animal endowed with plastic neurons; and some such systems may be consolidated by use whereas others can be weakened by disuse: see Figure 1.1. This is the *use–disuse* hypothesis, proposed by Tanzi, adopted by Ramón y Cajal, and refined and exploited by Hebb in his classical work *The Organization of Behavior* (1949). In another seminal work, *A Theory of Intelligent Behavior* (1976), Bindra proposed and applied a substantial modification of Hebb's idea of a neuron assembly: the psychons (which he called *pexgos*) need not have a fixed spatial location but may be itinerant, so that every mental act, even a repetitive one, may consist in the formation of a fresh psychon.

There is considerable empirical evidence for the hypothesis that neurons assemble into multineuronal systems, which in turn are organized into supersystems (Mountcastle, 1978; Szentágothai, 1978). There is also a fair amount of evidence for the hypothesis of plasticity (cf. Flohr and Precht (Eds.), 1981) and, in particular, for the hypothesis that "use" strengthens synaptic transmission. (See, e.g. Bliss (1979) and Goddard (1980) for both short term and long term changes in connectivity.) Finally, there is mounting evidence for the hypothesis that mental processes are processes in plastic neural systems (e.g. Baranyi and Fehér, 1981). All such

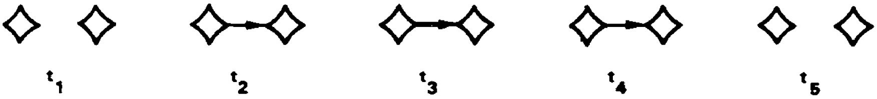


Fig. 1.1. Use–disuse. At time t_1 the two neurons are disconnected from one another. At t_2 one of them synapses onto the other: a system has been formed that may do things neither of the isolated neurons can. By t_3 the connection has been strengthened by use. At t_4 the coupling has weakened through disuse. At t_5 the connection has ceased through prolonged disuse or degeneration (e.g. pruning of dendrites).

hypotheses and data suggest a theoretical framework possessing considerable explanatory and heuristic power (Bunge, 1980a). However, this is just a beginning: we need plenty more detailed neural models, at least one for every cognitive function, and precise tests must be devised and conducted to check them. In any event, if we wish to understand the mind, and in particular cognition, we must give up the traditional arrogant armchair attitude and study the body, for minding is a brain process.

(Several mathematical models of learning psychons incorporate Hebb's use and disuse hypothesis. For example, Amari and Takeuchi (1978) postulate that the learning or self-organizing process consists in a change of the synaptic weights w_i associated to the input signals x_i according to the rate law

$$\tau \dot{w}_i = -w_i + cx_i H(w \cdot x - h),$$

where τ is a time constant, c a positive number measuring the learning efficiency, H the unit step function, h the threshold value, and $w \cdot x$ the scalar product of the synaptic weight vector w by the input vector x. If any of the stimuli are inhibitory, the corresponding synaptic weights are negative. By solving the above system of equations one finds that $w(t)$ is the weighted time average of $H(u)x(u)$ from $u = -\infty$ to $u = t$ with the weighting factor $(1/\tau)\exp[-(t-u)/\tau]$. By assuming that x is a random sequence of a certain type, a number of interesting consequences follow. But, because the neurons in this model are not supposed to fire unless they are excited externally, they are capable of learning only what they are taught. Plasticity, then, is not enough to account for spontaneous learning: if we wish to represent the creativity characteristic of conceptual learning we must include spontaneous neural activity. One way of doing so is to replace the above first order differential equation with a 2nd order one.)

In sum, we assume that the nervous systems of animals of certain species have plastic subsystems, i.e. neural systems with modifiable connectivity, as well as rigid neural systems. We also assume that every mental function is a process in some plastic neural system. For example the formation of a polysyllabic word, or of a sentence, of the form *abc*, probably consists in the sequential activation of neural units A, B, C, in the same order. This speculative hypothesis explains also thought and speech anomalies such as spoonerisms. Thus saying 'pointcounter' when intending to say 'counterpoint' might be due to the wrong timing of the psychons engaged in the production of 'counter' and 'point'. And in turn this time reversal might be caused by the temporary inhibition, or else the delay, of the first psychon, which inhibition (or delay) could be due to a number of factors, such as a strong coupling between the delayed psychon and a third psychon, or to external interference.

Our hypothesis that the nervous system is the organ of cognition is of course the very first postulate of any biologically oriented psychology and epistemology. This view is at variance with the fashionable cognitivist school and its computational or "functionalist" philosophy of mind (Putnam, 1960; Fodor, 1975, 1981; Pylyshyn, 1980). These are the basic tenets of cognitivism:

(i) Cognition is computation, and as such it is not a prerogative of brains: it can also be performed by computers. (Yet many cognitive processes, in particular perception and imagination, are not "intentional rule-governed phenomena" but occur spontaneously and unconsciously.)

(ii) Computation takes place in the mind, which is a kind of computer. (By pushing computation into an immaterial mind, the study of it does not fall under the jurisdiction of natural science. The fact that some computations can be simulated on computers does not prove that computation *in vivo* is essentially the same as its artificial mimicry.)

(iii) Any suitable theory of cognition is basically a theory of Turing machines. (This theory is so extremely simple that it cannot account for the humblest thing in the universe, such as an electron or a photon—let alone for a brain with its myriads of neurons and their plastic connections.)

(iv) Knowledge of the brain is irrelevant to understanding cognition and, in general, matter does not matter to mind. (Since cognitive processes are brain processes, neuroscience is essential to the advancement of cognitive psychology.)

(v) Artificial intelligence is the proper foundation for cognitive psychology, not the other way round. (So far this approach has produced only analogies. It is to be expected that, as neuropsychology advances, it will be able to guide work in artificial intelligence for, after all, imitation presupposes knowledge of the real thing.)

For the reasons advanced in parentheses, we regard cognitivist psychology not only as mistaken but as wrong-headed and paralyzing: as a collection of simplistic and superficial metaphors, a modern version of psychoneural dualism resulting in a further estrangement of psychology and epistemology from science, in particular biology. (For further criticisms see Bunge (1980a) and Vol. 6.)

1.2. *Brain States and Processes*

We have assumed that *every cognitive process is a process in some plastic subsystem of the nervous system of an animal.* Consequently an animal lacking a nervous system—be it small like an amoeba or large like a sponge—cannot know anything. It can of course react to stimuli of certain kinds and even act spontaneously—but so can any lump of matter.

We also assume that the nervous system, far from being a mushy thing, is a highly organized biosystem composed of subsystems each of which discharges certain peculiar functions in addition to the functions common to all living systems (such as metabolism and protein synthesis). Sometimes the specialization goes down to the individual neuron level. Thus the neurons in the mammalian visual cortex are so specialized that, whereas some of them detect only vertical lines, others react only to lines tilted at a 20° angle, and so on; still others detect short lines, wide lines, or what have you. (See Hubel and Wiesel, 1977; Gilbert and Wiesel, 1979.) In short the individual neurons of that kind analyze the incoming sensory inputs simply by selectively detecting some of their features. The integration or synthesis of these specialized activities is done by multineuronal systems. Traditional neuropsychology held that the brain is a sensory analyzer: we know now that it performs both the analysis and the synthesis of incoming stimuli, and also that it acts as a shock absorber for most of them, particularly the repetitive signals.

Whereas some cognitive processes occur in rigid (committed, wired-in, programmed) neural systems, others engage plastic neural systems as well, or perhaps exclusively. Thus the mere detection of ordinary stimuli requires only rigid neural systems, and so does the classing of stimuli into harmful, harmless, and beneficial. On the other hand locations or times, faces or letters, tricks or skills of any kind, presumably take plastic neural systems or psychons. The latter have been experimentally identified in a few cases. Thus Lindauer (1971) trained bees to forage at certain places, and then transplanted their tiny mushroom bodies into the brains of untrained bees. After recovery the latter made a bee line for the feeders. Knowledge can then be surgically removed and transplanted into a different animal—much to the chagrin of the mentalist.

The hypothesis that every cognitive activity is a process in a specialized neural system (i.e. the specific activity of it) can be elucidated as follows. To say that thing x is *active* at time t is the same as saying that at least one property of x is changing at t. Now, every property is lawfully related to at least one other property of the same thing: i.e. there are no stray properties. (For this ontological principle see Vol. 3, Ch. 2, Section 3.3.) So, if one property of a thing changes, other properties of the same thing will change as well: i.e. every change is multidimensional. For this reason we should always consider clusters of properties and, in principle, all the properties of a thing. If we do so we can speak of the states and the changes of state of the thing of interest. Every *state* is a list of properties, and every change of state of a thing is called an *event* occurring in the thing. The sequence or list of all the changes undergone by the thing during a time interval is called the *process* occurring in the thing over that interval. (There are no events or processes in themselves: every change is a change in some thing. Nonthings, such as immaterial souls and disembodied ideas, are in no state and *a fortiori* cannot change state.)

(A general and useful way of exactifying these ideas is as follows. First, represent each property of a thing of kind K as a mathematical function, in principle a time dependent one. (Never mind the other dependent variables for the moment.) Call F_K^i the function representing the ith property of thing of kind K. Next, collect all these functions into a list or n-tuple $F_K = \langle F_K^1, F_K^2, \ldots, F_K^n \rangle$, called the *state function* of an arbitrary member of K. The value $F_K(t)$ of the state function at time t may be pictured as an arrow representing the *state* of the thing at time t. As time goes by the tip of the

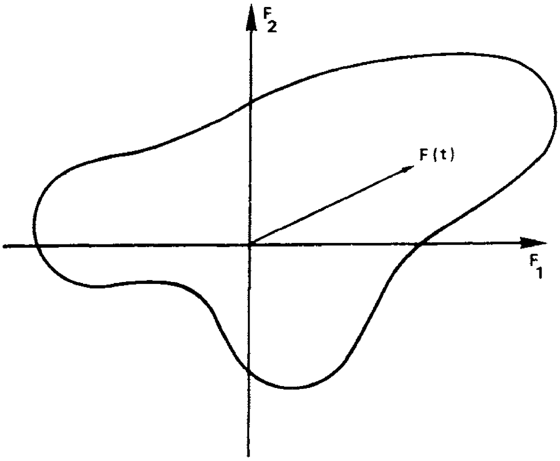


Fig. 1.2. State space for an (imaginary) thing characterized by two properties, represented by functions F_1 and F_2, combining into the state function $F = \langle F_1, F_2 \rangle$. The tip $F(t)$ of the "vector" F at time t represents the state of a thing of the given kind at t.

"vector" $F_K(t)$ spans a region of an n-dimensional space called the *state space* for things of kind K. (See Figure 1.2 and, for details, Bunge (1977a) or Bunge (1979a).) In other words, the "life history" of a thing of kind K is representable as a trajectory (or a bundle of trajectories) in the state space $S_K = \{F_K(t) | t \in T\}$, where T, a subset of the real line, is the set of all instants. The laws of things of kind K restrict the possible values of the state function, and therefore limit the state space for K's to a subset of the set of all logically possible states.)

(A state space for a species K represents all the lawful states of unspecified members of K. To individualize any state of a particular member of K we must specify its particular constraints and circumstances. Call F_K^a the state function representing an individual a of kind K. Knowing all the values of a particular F_K^a over a certain time interval amounts to knowing the history of the individual concerned during that interval. More precisely, we may say that the *history* of an individual of kind K over the time interval τ is representable as the graph of its F_K^a during τ. See Figure 1.3.)

(Now, the history of a thing may be eventful or nearly uneventful during a given interval. I.e., the state function F_K^a for a given individual a of kind K may change considerably or only negligibly over a given period. If the history is eventful we call it the *activity* (or process) of the thing concerned. Speaking formally, the activity of, or process in thing a during period τ, is

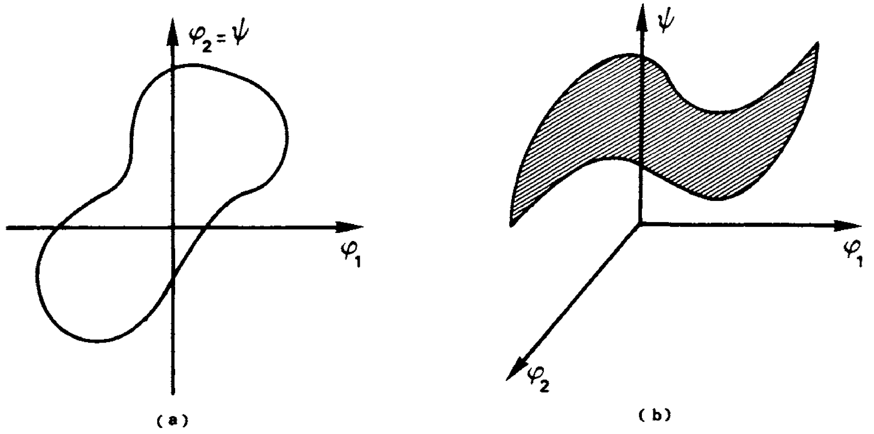


Fig. 1.3. Projection of state space of a plastic neural system. In (*a*) one of the biological properties (axes) is identical with a psychological one. In (*b*) the psychological property is some function of two biological properties—but is itself a biological property.

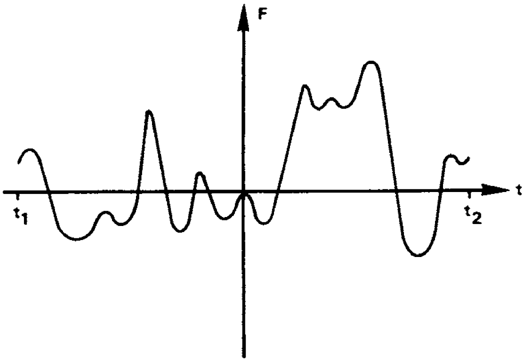


Fig. 1.4. History of a thing during the time interval $[t_1, t_2]$.

representable by the set of ordered pairs $\langle t, F^a_K(t)\rangle$ for t in τ, such that the rate of change $\dot{F}^a_K$ differs from zero:

$$\pi(a,\tau) = \{\langle t, F^a_K(t)\rangle \mid t \in \tau \ \ \& \ \ \dot{F}^a_K(t) \neq 0\}.$$

This graph represents the *total* activity of a over τ. See Figure 1.4.)

(In order to represent only the peculiar or specific activity of a thing we must subtract whatever other things do over the same period. Since we are interested in the specific activities of certain subsystems of an entire animal, we shall take "the other things" to be the parts of the animal other than the system of interest. To indicate that x is a *part* of b we write '$x \sqsubset b$' (Vol. 3, Ch. 1, Section 1). Thus the *specific* or *peculiar* process (or *activity* or *history*) going on in subsystem a of animal b during τ is representable by

$$\pi_s(a,\tau) = \pi(a,\tau) - \bigcup_{x \sqsubset b} \pi(x,\tau), \quad \text{with} \quad x \neq a,$$

where, for any sets A and B, $A - B = A \cap B$, '$\cap$' stands for intersection and '$\bar{B}$' for the complement of B. In other words, the specific function (or process) of system (e.g. organ) a in animal b equals its total function minus the total functions of all the remaining parts of the same animal. Thus imagine a system with only two components, and call them a and b. Suppose that their respective activities or processes during a given time interval are $\pi(a) = \pi \cup \pi_1$ and $\pi(b) = \pi \cup \pi_2$. Obviously their specific activities are $\pi_s(a) = \pi(a) - \pi(b) = \pi_1$, and $\pi_s(b) = \pi(b) - \pi(a) = \pi_2$.)

For example, seeing is the specific activity of the visual system, which in the higher vertebrates includes the visual cortex and is coupled to the motor system. (Every sensory system is a subsystem of its corresponding sensory-motor system. Thus looking is accompanied by saccadic movements, sniffing by puckering the nose, tasting by moving the tongue, listening by turning the head, and touching by moving a limb. After all, the sensory stripe is right next to the motor stripe in the cerebral cortex. In short, there is no perceptual learning without motor activity. In other words perceptual learning is not just a matter of receiving passively sensory inputs, but of combining a selection of such inputs with motor outputs: see Held and Hein (1963).) Likewise making a computation is presumably the specific activity of some psychon(s) in the association cortex; becoming conscious of one's own thoughts or feelings is presumably the specific activity of a neural system monitoring other neural components of the same brain, and so on.

Neuroscientists and psychobiologists are working on the problem of mapping the brain, in particular the cerebral cortex—i.e. of finding out what (neural system) does what (function or activity). This task was started by Penfield (1950) who, by stimulating electrically certain points of the cerebral cortex of waking patients, evoked in them various perceptions, images, memories, and thoughts. A suitable map of the brain would equate every cognitive function with a particular activity of some neural system or other (cf. Olds, 1975). Note that we have written *equate* not just *correlate* each cognitive function with a neural activity. Whereas the former expression suggests the identity of cognitive function and specific brain function, the latter expression suggests only a vague association between cognition and brain functioning in the style of psychophysical parallelism or interactionism. The dualist, to whom the mind is not a collection of brain functions but a separate substance, has no real use for brain mapping.

This mapping is such a formidable enterprise that it may be doubted whether it will ever be completed (Hebb, 1980). Consider: the plastic part of an individual human brain is composed of at least 1,000 million neurons. If each psychon has between 100 and 10,000 neurons, the total number of psychons is at least one million—and many times this figure if psychons are itinerant rather than fixed. Moreover it is well known that there are considerable differences among individual brains, so that the precise location of a particular cognitive function in one person is at most a very rough indication for its location in the next person. Nevertheless there has been considerable progress in this area over the past few years, as new techniques of imaging brain processes have been developed. These techniques have laid to rest the holistic hypothesis that the brain functions either as a totality or not at all.

TABLE 1.1
Cognitive deficits or disorders and their plausible neurophysiological explanation

Name or description	*Plausible explanation*
Blindness	Visual system damage or malfunction
Impaired vision	Multiple sclerosis, brain tumor, or dioptric defect
Visual agnosia	Lesion in occipital or posterior temporal regions or left hemisphere
Visual illusions and hallucinations	Lesions at any level in the visual system from eye to cortex
Deafness	Auditory system damage or malfunction
Unawareness of left half of body	Right parietal lobe damage or malfunction
Loss of attention	Reticular activation system or thalamus damage or malfunction
Amnesia	Bilateral hippocampal or diencephalic damage or other lesions or diseases
Loss of short term memory	Hippocampus damage or malfunction
Loss of curiosity	Frontal lobe damage
Loss of consciousness	Reticular activating system damage or malfunction
Aphasia	Wernicke or Broca area damage or malfunction
Dyslexia	Gyrus lesion
Agraphia	Temporal–parietal occipital lobe damage, retrorolandic lesion
Acalculia	Temporal or parietal lobe damage

Table 1.1 lists some of the cognitive deficits or disorders most frequently "seen" by clinical psychologists, psychiatrists, neurologists, and neurosurgeons. (See Pincus and Tucker, 1978; Hécaen and Albert, 1978.) We seem to know more about brain abnormality than about normality for two reasons: (*a*) the study of malfunction is one of the main ways of finding out what the normal function of a component is: we notice the light bulb in the dark staircase when it has burned out; (*b*) there are more researchers and means engaged in medical research than in neurobiological research—because of a science policy inspired in a pragmatist philosophy of science and technology.

Many workers speak of *information processing* where we speak of *brain processes*. That mode of speech is justified insofar as the nervous system does possess components capable of detecting, absorbing, transducing, or propagating signals throughout the system. However, it does not follow that the statistical theory of information is actually used to advantage to describe or explain such processes, and so no specific neural models need be constructed and tested. In fact information theory (Shannon and Weaver, 1949) applies only to messages transmitted through noisy channels and having definite probabilities of being received and decoded, whereas neuroscience and psychology are interested in *specific* biophysical, biochemical and biological processes occurring in the nervous system. When neuroscientists say that '*A* sends information to *B*', or that 'There is a flow of information from *A* to *B*', they ought to say simply that *A acts on B through C*. And when they write '*B* processes information coming from *A*' they ought to say simply: *A acts upon B*, triggering in *B* a process peculiar to *B* and different from any processes in *A*. After all they are dealing with systems of a very special kind, namely nervous systems, not—like the information theorist—with all kinds of information processors.

Besides, the brain is not just a processor of incoming stimuli: it also generates signals. True, S-R (or behaviorist) psychology, as well as "ecological" psychology (Gibson, 1979), hold that the brain does not create anything but picks up information from its environment. But of course any neuroscientist knows that the brain of the higher vertebrate has a constant autonomous activity. And any psychologist might argue that the view that there is no creation of new information leads to infinite regress. Indeed, if all I know is what I learned from *A*, who in turn learned it from *B*, and so on, then the only way of putting an end to the chain is to postulate either a Prime Mover or a Platonic realm of eternal ideas, neither of which is a useful and testable hypothesis. Besides, we should recall that, although the environment of mankind has always contained electromagnetic fields, molecules, and cells, it took millions of years of biological and cultural evolution for the emergence of organisms capable of creating the corresponding ideas. In short, information is not out there for grabs: the organism absorbs most of the stimuli impinging upon it, transforms a few of them into processes of other kinds (i.e. transduces them), and generates its own, part of which it transmits to other animals.

2. COGNITIVE FUNCTIONS

2.1. *Perceiving, Thinking, and All That*

We get to know about the world and ourselves through perception and ideation, two distinct but intimately related neural processes. Perception is the last stage of a process that normally starts as mere detection or sensation in some neurosensor (or neurodetector), such as a nerve ending in the skin or in the tip of the tongue. Sensation is the specific activity of sensory systems such as the olfactory and the auditory systems. The neuroreceptors are quite unlike physical or chemical receptors, in that they are under the constant "downstream" action of the central nervous system. Consequently the state of a sensory system depends on the state and history of the brain as well as on the nature and intensity of the stimulation. Hence we are unlikely ever to sense the same stimulus twice in exactly the same manner. Yet we may perceive the two stimuli as identical even if they are not. Moreover we may miss a tone or a motion altogether unless we were expecting it. (See Figure 1.5.) To put it into other words: the response is a function of both stimulus and internal state of the organism. (When Hunter said this in 1929 he was not listened to: the behaviorists were prepared to hear only results fitting the stimulus-response model, which ignores the state of the organism.)

In the higher vertebrates sensation is processed in the sensory cortical "area", which is divided into the primary, secondary, and tertiary "areas". The primary "area" is a component of the corresponding sensory system, and apparently it is particularly malleable during the early stages of development, which is when the animal learns to smell, see, hear, etc. We shall assume (Hebb, 1966; Bindra, 1976) that, whereas sensation is the specific activity of a sensory system including the primary sensory cortical "area", perception is this activity together with the specific activity of the plastic neural systems directly attached to it, i.e. the secondary and tertiary

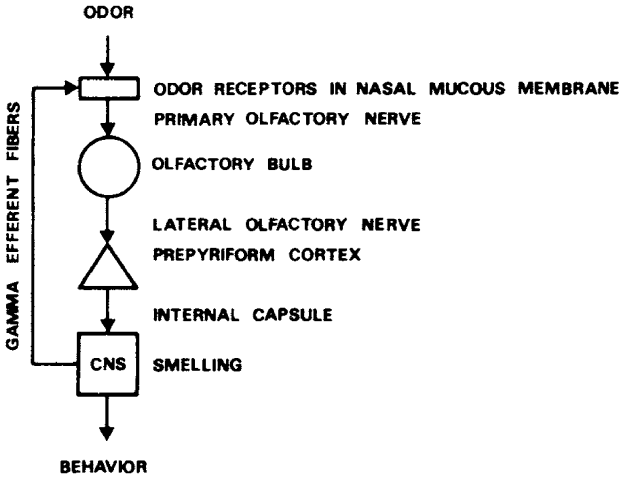

ODOR
ODOR RECEPTORS IN NASAL MUCOUS MEMBRANE
PRIMARY OLFACTORY NERVE
OLFACTORY BULB
LATERAL OLFACTORY NERVE
PREPYRIFORM CORTEX
INTERNAL CAPSULE
CNS
SMELLING
GAMMA EFFERENT FIBERS
BEHAVIOR

Fig. 1.5. Smelling, a primary cognitive activity, as a nervous system process. The diagram sketches the large scale organization of the olfactory system in mammals. Note the distance between odor reception and odor perception (smelling). Note also the feedback of the CNS (central nervous system) on the receptors.

"areas", as well as whatever other plastic neural systems may be activated by them. Thus there is a long neural way between odor (or light or sound or pressure) reception and its perception, i.e. smelling (or seeing, hearing, or feeling).

Far from being isolated from other neural activities, perception is linked with a number of them, including ideation and behavior. (See Figure 1.6.) This link accounts for the fact that perception can occur in the presence of fragmented stimuli, on the basis of minimal cues, and even in the absence of sensory stimulation. It also accounts for the various constancies (of, e.g. distances and shapes) as well as for spontaneous perceptual processes such as hallucinations and those involved in dreams. This view is at variance with the causal theory of perception, according to which every percept is caused by an external thing, and this in an invariable manner. This view, which is still going strong, ignores the ongoing spontaneous activity of the brain as well as the complexity of the processing of signals coming from the sensors. Perceiving is never a passive effect of the thing perceived: it is not copying·but constructing (Neisser, 1967; Bindra, 1976), and therefore it is midway between sensing and thinking (Gregory, 1970). When perceiving something we construct a percept of it with sensations, memories, and expectations.

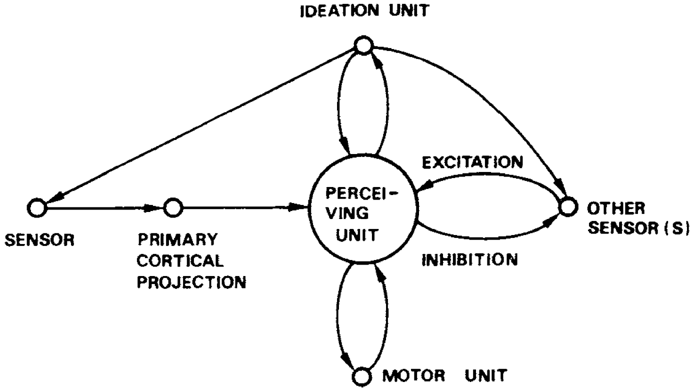


Fig. 1.6. Perception is linked to sensation, ideation, and behavior. Schema inspired by Bindra (1976), who in turn modified it from Hebb (1949).

In other words, the stimuli coming from the perceived object elicit *changes in the ongoing activity* of the nervous system rather than causing the latter. (This is why we assume that an animal perceives an external object as the symmetric difference between the activity of the corresponding process in the presence of the object, and in the absence of it: see Bunge, 1980a, Postulate 4.1.) Therefore there can be perception without external stimuli, as in hallucinating and dreaming, but not without a perceptual system. It follows that there can be no extrasensory perception (ESP). And if there is no ESP (clairvoyance, telepathy, precognition, and the like), then parapsychology is an illusion (Wundt, 1879; Hansel, 1980). What remains to be explained is not the alleged facts of parapsychology but the gullibility of parapsychologists—a task for social psychologists (Alcock, 1981).

Strictly speaking we do not perceive things but events (changes) occurring in things. Thus we cannot see an object unless it emits or reflects photons of a certain energy. Moreover some such events, to be perceptible, must trigger processes ending up in some neurosensor or other. Unless the two conditions are jointly satisfied the object remains invisible. And the way the animal perceives the events occurring in its neurosensors is by *mapping* them into events in some of its perceptual systems. Admittedly this mapping is partial (not everything gets represented) as well as distorted. Nevertheless it is a genuine map, i.e. a representation of certain sets (of events) into another set (of events). The first set (or domain of the function) is a collection of events in the environment or the periphery of the animal, the second set (or codomain) is a collection of events in the perceptual

$$\begin{tikzcd}
\text{EXTERNAL EVENTS} & 2^{E(e)} \arrow[r, "\text{STIMULI}"] \arrow[dr, "\text{EXTERNAL WORLD MAP}"'] & S(b) \times 2^{E(b)} \arrow[d] & \text{BODILY EVENTS} \\
 & & 2^{E(c)} & \text{PERCEPTUAL EVENTS}
\end{tikzcd}$$

Fig. 1.7. Events in the environment e of animal b are projected into pairs ⟨bodily state, set of bodily events⟩. This map dovetails with the next, or somatosensory map, which projects external events into cortical events, or points in the secondary or tertiary sensory area. The composition of the two maps yields the external world map for the corresponding sensory system. $E(e)$ is the set of possible events in the environment of the animal; $S(b)$ the state space of the animal and $E(b) \subset S(b) \times S(b)$ its events space. $2^{E(b)}$ is the family of all subsets of $E(b)$; $2^{E(c)}$ is the family of all subsets of the event space $E(c)$ of the cortical system concerned. Details in Bunge (1980a).

system of the animal. In every case the map or function depends not only upon the sensed stimuli but also upon the state of the animal. See Figure 1.7. Maps or internal representations of this kind allow the animal to "navigate" in its environment.

Whereas some memories are remembered, others are forgotten, temporarily or permanently. And those that are recalled are so, more often than not, in a distorted way: like perceiving, memorizing is creative—i.e. not just storing and retrieving but reconstructing (Bartlett, 1932). Donald Hebb used to ask his students to remember breakfast. They all "saw" themselves (with their mind's eye, as we say in ordinary language) sitting at table—an obvious construction. This constructive or creative feature of memory is ignored by the computer (or information-theoretic) metaphor, according to which the learned items are encoded, stored, and then retrieved, possibly with loss, never with gain. This holds literally for computer memories and library systems, but is grossly off the mark in the case of brains.

Since the storage models of memory are inadequate, some dynamical model is likely to prevail. One such model involves the following hypothesis: "Remembering consists in the fresh production or reconstruction of an item, not in the *re*-activation of some permanent representation of the item as such, and what item is remembered—the response production—is the pexgo [psychon] selectively activated at the time of remembering" (Bindra, 1976, p. 330). Rather than permanent changes in circuitry, there may be a propensity or disposition for certain neural systems to be activated, either spontaneously or induced by others. Such dispositions may be quantitated as the probabilities of certain synaptic connections. Any such probability increases with use and decreases with disuse or abuse (e.g. drugs), or just age. The ability to memorize is of course necessary for learning anything: patients who have lost their short term memory (as a result e.g. of brain injury) are incapable of learning anything but for some elementary behavior patterns. (Amnesic patients can be conditioned even though they fail to remember the conditioning process: Weiskrantz and Warrington (1979).)

Of all the brain functions thought is the most remarkable and the least understood. Forming concepts, transforming and interrelating them to construct propositions, questions and directions, are examples of thinking. (Many psychologists include visual images among thoughts—but then why not auditory, olfactory, and tactual images? We regard visual imagination as perception in the absence of sensory stimulation, and submit that it can help or hinder thinking, not however take its place.) Let us deal quickly with the formation of concepts and propositions.

One of the characteristics of the formation of concepts, as well as of the production of the corresponding verbal expressions, is that they may occur in the absence of external stimulation: they can be purely internal processes without sensory inputs or behavioral (motor) outputs. And, when elicited by external stimuli, concept formation (and the corresponding speech production) may involve lack of discrimination. Thus we can form the general concept of a tree perhaps because we possess plastic neural systems the specific function or activity of which is roughly the same whether they are activated by seeing or remembering a maple tree, a fir tree, or a birch tree: such psychons are generalists rather than specialists (Bunge, 1980a). An alternative account of concept formation is this. Every time an animal perceives a tree, several million neurons are excited, perhaps a different set every time. But there may be a thousand neurons common to all those different excitations. These will become organized into a neuron assembly, for neurons that fire together tend to stay together. The activity of this neuron assembly is (identical with) the general idea of a tree (Hebb, 1949). It is up to experiment to determine whether one of these hypotheses, or a third one, is true.

We hypothesize that forming complex concepts, propositions, problems, instructions, and other composite ideas consists either in psychon grouping or in sequential activation of two or more psychons (possibly cortical columns). Thus the concept of an even number would be formed by the pairing of psychons for "number" and for "even". An alternative hypothesis is that such a concept consists in the successive activation of the simple psychons concerned—an extremely fast succession indeed. (In this case language would certainly make a difference for, whereas in English and German 'even' comes before 'number', in the Romance languages it is the other way round.) However, the two hypotheses are mutually compatible. Indeed, psychon pairing may occur during learning (or relearning), whereas sequential activation may happen while using or applying the learned item, since language may be regarded as a vast storehouse of ready-made thinking bricks. The young child has to think hard to add two and three: we just use the result, encoded in a linguistic entity, when needed. (More on language and thought in Ch. 3, Section 3.2.) Once the learning period is over we can perform uncounted mental operations in a thoughtless ("mechanical") manner, just as we walk or drive relying mostly on automatisms. But in either case we may have to resort to new thinking when faced with novelty. (Unfortunately we do not always do so: witness the ideologue who, because he cannot cope with novelty, wants the world to stand still.)

Psychon pairing and sequential activation explain much but not everything. In particular they do not account for our occasional ability to direct thought or "concentrate" it on one task in spite of external (sensory) and internal (spontaneous) excitations that could be diverting. However, the directedness of thought is explained psychologically by the fact that it always occurs in a context: that of the problem at hand (or rather brain), which in turn is embedded in a body of knowledge. In turn, this context has a possible physiological explanation: it could be the background activity of psychons associated with those doing the thinking (Hebb, 1976). And the fact that we are able to restrict the activity to some psychons (rather than engaging the entire plastic brain) can be explained in terms of lateral inhibition, or the containment of the spreading of nervous excitation. To be sure all this is speculation, but reasonable and fertile: hypotheses such as these underlie empirical research into the physiology of thinking. (See Bechtereva, 1978; Goddard, 1980.)

In any event we assume that a thought, be it a concept, proposition, inference, problem, or command, is a brain process. More particularly, we postulate that every thought is a specific process occurring in some plastic neural system or psychon of some animal. Hence a thought may be described as the sequence of states of such psychon over the time lapse it takes to be completed. The particularities of such process are likely to vary from individual to individual, and even from moment to moment in one and the same individual. Thus it may well be that nobody undergoes exactly the same brain processes when thinking of the number 5 at different times: it could be that different psychons are activated at different times (Bindra, 1976). However, any such thoughts of the same construct must be equivalent in an important respect, i.e. the processes must have the same general shape, for otherwise they would not consist in thinking of the number 5.

(Suppose neuropsychology discovers the exact nature of such equivalence, and call it $\sim$. Further, call $\theta_{ij} = \pi_s(b_i, \tau_j)$ the specific process occurring in brain b_i, during period τ_j, when thinking of a certain construct. Since all the processes consisting of thinking of the same construct are $\sim$-equivalent, we write: $\theta_{ij} \sim \theta_{mn}$ for all i, j, m, n. We now form the equivalence class (set of all $\sim$-equivalent processes) $[\theta_{ij}]_\sim$. Finally, we identify the construct c with this equivalence class, i.e. we set

$$c = [\theta_{ij}]_\sim.$$

In other words, *a construct is not an individual brain process but an equivalence class of brain processes* occurring in different brains or in the same brain at different times (Bunge, 1981a). As long as the exact nature of the equivalence relation $\sim$ is not defined, this is a programmatic or heuristic hypothesis that can guide the very research into the identification of $\sim$. More on such equivalence in Ch. 2, Section 1.1.)

The psychobiological conception of thinking explains the association of ideas, that mainstay of empiricist psychology and epistemology, as well as its *bête noire*, intellectual creativity. Association is explained as the interconnection of psychons, and originality as the formation of radically new psychons. Neither is mysterious: interconnection because the whole brain is a system, so that anything happening in one of its components influences some other components. (The influence may be positive and may get established, as in the case of association, or it may be negative or inhibiting, as when every time we think of mathematics we tend not to think of sports.) And originality is not mysterious because of the remarkable spontaneity and plasticity of the cortical neurons.

(The ideas of association and of originality may be formalized roughly as follows. Let S_i be the collection of cognitive acts that subject i performs up until a certain time: call it the *cognitive space* of i. Further, call x and y any two members of S_i, and interpret their concatenation $x \circ y$ as "cognitive act

x associates with (or is followed by) cognitive act y". Finally, call 0 the act that does not increase knowledge even if it was intended to do so; more precisely, 0 represents an arbitrary unsuccessful cognitive act—not to be mistaken for $\varnothing$, or total ignorance. We postulate that the structure $\langle S_i, \circ, 0 \rangle$ is a noncommutative monoid. This amounts to the hypothesis that there are no isolated cognitive acts, but any cognitive act may be associated with some other cognitive act. And we may stipulate that a cognitive act is *original* with subject i if, and only if, it belongs in S_i but not in the cognitive space of any other animal up to the given time.)

2.2. *Learning*

Learning can be conceived broadly or narrowly: as adaptation or as neural modification, temporary or permanent. In the former case one may say that an organism which has developed immunity to some antigen has "learned" to face the latter, and even that a population which has become adapted to a changed environmental—after a wasteful selection process—has "learned" to live in it. Although such a broad construal of learning puts it in a biological perspective, it erases the specificity of the individual neural process accompanying the acceptance or generation of new information. (The same charge can be laid against the information processing view.) Therefore we adopt the narrow construal of learning as neural modification.

We assume, then, that an animal *learns* something if it forms at least one new psychon (plastic neural system) or combines in a novel fashion two or more psychons that had been established previously. And we call *knowledge* of an animal at a given time *the set of all the items it has learned* up until that time—i.e. the collection of changes in its plastic neural supersystem. (Such changes need not all consist in permanent alterations of connectivity: some may consist in changes in the probabilities of certain synaptic connections.) Needless to say, accumulating experience (learning) is not passive recording but a process where signals are selected (mostly filtered out) and transformed—distorted and enriched. In short, learning is *creative* even when it consists only in learning what other animals have learned before. It is creative in the sense that it is not automatic and that it is new to the individual.

Like any other brain process, learning can be represented as a change in an abstract space representing the possible (lawful) states of the brain. An animal engaged in a learning process can be said to expand its plastic neural

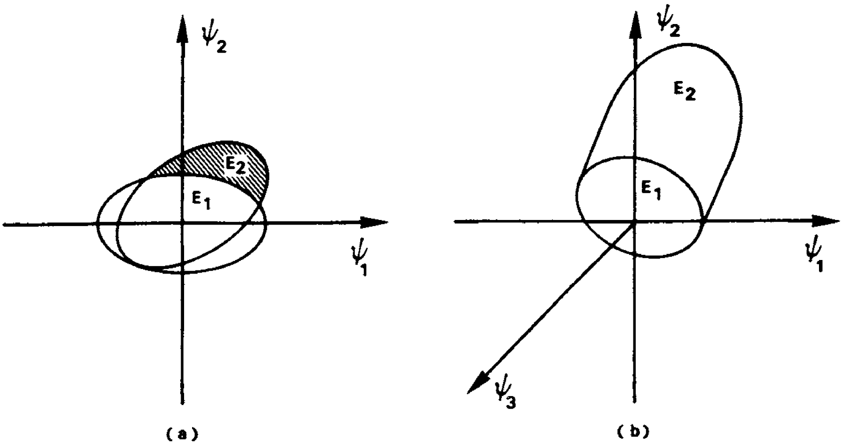


Fig. 1.8. Two kinds of learning. (*a*) Learning of one more item of a given kind (with eventual loss of some experience). (*b*) Learning something qualitatively different is represented by the intervention of a third coordinate axis in the neural state space. Adapted from Bunge (1980a).

state space. This extension can be quantitative or qualitative: the former when the animal learns one more trick of a kind it has sampled before, and the latter when it learns something of a different kind. Thus learning one more word and one more algorithm fall into the first category. On the other hand the learner's neural state space expands qualitatively—i.e. new axes sprout—when it masters a new subject for the first time. The most rewarding aspect of research and advanced teaching is that it encourages the sprouting of new axes in one's brain state space. See Figure 1.8.

Several learning theories are compatible with both the neuroscientific and the psychological data. Only a few of them are deep, i.e. concerned with changes in neural systems, but they are not detailed enough. Most learning theories are purely descriptive: they disregard brain mechanisms and concentrate on the necessary and sufficient environmental conditions for learning. (In particular, the fashionable information-processing view, which ignores the nervous system, is a metaphorical description of what happens during the learning process.) However, some of these theories do capture certain important features of learning, particularly that it is a stochastic process, i.e. a sequence of events every one of which has a definite probability.

Figure 1.9 illustrates the class of stochastic learning theories in the case of single trial learning. The animal is assumed to have probability p of

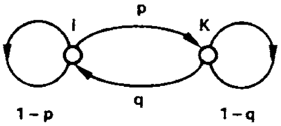


Fig. 1.9. A stochastic model of all-or-none learning events. The arrow from *I* to *K* symbolizes learning, and the converse forgetting. The closed loops represent not learning (left) and not forgetting (right).

learning a given item at a trial, hence probability $1 - p$ of not learning it; and it has probability q of forgetting it, *ergo* $1 - q$ of remembering it. If multiple trials are allowed a well known pattern emerges. For any given task the number of successes increases with practice (i.e. number of trials) as well as with the number of successes. (Successful learning is its own reward.) Likewise the number of errors decreases with practice as well as with the number of errors that get corrected. (Otherwise the process is haphazard, i.e. no trend emerges.) The result is the logistic curve: Figure 1.10. This curve summarizes a purely descriptive theory of learning that does not account for the learning mechanism—let alone its neurophysiological "substrate" or "correlate". It is a datum to be explained by neurophysiology.

Another view of learning that fails to make contact with neurophysiology is innatism, according to which "Learning is primarily a matter of filling in detail within a structure that is innate" (Chomsky, 1975, p. 39). On this view knowledge of a certain kind, i.e. pre-existing—in

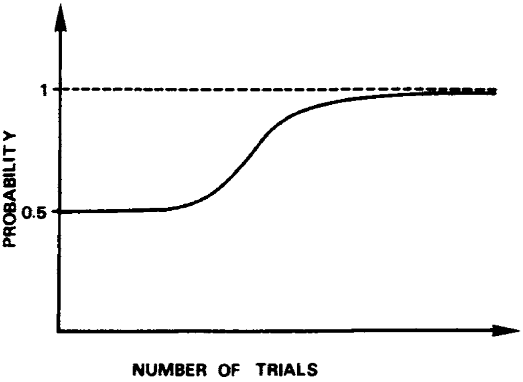


Fig. 1.10. The logistic learning curve: if errors are corrected, the probability of correct performance approaches 1 as the number of trials increases.

particular innate—knowledge, makes learning possible, not the other way round (Chomsky, *op. cit.*, p. 118). Needless to say, there is not the slightest shred of empirical evidence for this speculation. On the contrary, the anatomy, physiology and psychology of development show that the newborn human knows nothing, learns only gradually, and cannot skip any stages because learning keeps pace with the maturation of the nervous system.

The existing learning theories are nonspecific: they are assumed to represent motor learning (e.g. learning to ride a bicycle), perceptual learning (e.g. learning to discern a pattern in an assemblage of dots), conceptual learning (e.g. learning new concepts and propositions), etc. Yet it is obvious that there are big differences among these various kinds of learning, and it is well known that they are done with different subsystems of the brain. Thus conceptual learning, unlike learning of other kinds, involves understanding and it may proceed without any sensory-motor concomitants. So, in addition to general learning theories we need learning models representing each a given type of learning. For example, a model of motor skills learning, even if superficial (i.e. non-neurophysiological), should account for the interplay of behavior and cognition. Thus in the higher vertebrates stimuli that can teach (unlike those that elicit only reactions of the knee jerk type) do not trigger responses automatically but elicit the intervention of cognitive activities such as anticipation and recall.

In other words, the learning process initiated by an environmental stimulus is not summarizable by a stimulus-response arrow. Instead, what happens is that the stimulus is paired to some cognitive state, and both produce jointly a response. (In short, instead of $s \mapsto r$ we have $\langle s, c \rangle \mapsto r$.) In turn, the response has some consequence or outcome (internal or external). If the animal receives information about such outcome, this information may change its cognitive state, producing next time a different response, hence a different outcome, in the presence of the same stimulus. In short,

First trial ⟨stimulus, cognitive state⟩ ↦ response ↦ outcome.

Second trial ⟨stimulus, new cognitive state⟩ ↦ new response ↦ new outcome.

By comparing the successive outcomes the animal may learn whether they converge to the goal or desirable state. If they do not, it may correct the strategy—or give up. This sektch of a feedback learning model can be mathematized, "embodied" in a neural system, and in principle tested both behaviorally and neurophysiologically.

Not only learning theory but most of psychology and the whole of epistemology are attempts to answer the question 'How do we learn?' The empiricist answer is: 'I learn from my own experience and provided my behavior is reinforced (e.g. my ideas are confirmed)'. Skeptics and idealists, on the other hand, deny that we learn from experience—unless perhaps from negative experience, i.e. from mistakes. As is so often the case, each of the contenders holds a grain of truth. We do learn something (not everything) from experience—provided our nervous system is in a receptive condition, we are motivated (and thus willing to make an effort), and the material to be learned is somewhat organized.

That the nervous system must be in a receptive condition is obvious, although this point is ignored by both empiricists and idealists. No matter how well a learning material is presented, a person under the influence of alcohol or some other drug may be unable to learn it. But a healthy brain is not enough, nor is skillful teaching sufficient. Thus whereas Pavlov believed that we can learn automatically from paired experiences (e.g. sound of a bell and temporally contiguous presentation of food), it has been found that repeated paired experiences do not guarantee learning. (See Dickinson and Mackintosh, 1978.) It has been conjectured that, for somebody to learn to do B whenever A happens, he must somehow realize that A and B may be associated (Dawson and Furedy, 1976).

Learning requires also some motivation: an animal without motivation to learn will not explore (Bindra, 1974). Motivation for learning can be *intrinsic* (curiosity) or *extrinsic* (for the sake of a reward other than satisfying curiosity). Hence we can teach ourselves, and others can teach us, to regulate both motivation sources. Thus the standard teaching techniques, by emphasizing extrinsic reward (praise, prize, raise in status, etc.) may inhibit adventurousness and may favor a mercenary attitude towards learning. (Anyone trained to respond primarily to external rewards will say to himself that, since he wants a sticker or some other prize, he should pick problems he knows how to solve safely and quickly rather than problems that will keep him wandering and wondering.) In any event, learning is the easier the more strongly motivated—and motivation is in turn enhanced by learning.

Learning, then, is not just picking up information and processing it in a mechanical way: it is a highly selective and creative process (Piaget's *assimilation*). Nor does the brain remain unaffected by learning: it restructures itself as a result of learning (Piaget's *accommodation*). To be sure an isolated brain would be unable to learn anything; moreover it would not even function normally, as shown by the experiments on sensory deprivation. However, looking, listening, sniffing and touching are not enough to take cognizance of anything beyond appearances. Recall that humans have always reproduced, yet did not discover the reproduction mechanism until a couple of centuries ago; they have presumably been speaking for many thousands of years, yet linguistics was not born until the 19th century; they have always lived in the midst of a gravitational field, but this fact was not discovered until recently; and they have been playing games of chance for millennia, yet the earliest ideas about chance and probability were not conceived until the 17th century. In short, sense experience and action are poor teachers: most people do not learn readily from either. (It is easier to learn to perceive and do.)

This difficulty to learn from sensing and doing has multiple sources: (*a*) appearance, unlike reality, is poorly structured; (*b*) all signals are embedded in noise; (*c*) perception and memory are highly selective: whereas at times they discard details, at other times they pass over essentials; (*d*) motivation is often lacking; (*e*) learning much and fast requires lots of knowing; (*f*) we learn most easily those items that happen to be consistent with our belief system, and tend to ignore those which are dissonant—as is the case with unfavorable evidence for our pet beliefs; (*g*) we are seldom equipped with comprehensive and coherent theories helping us "make sense" of experienced events, so we have nowhere to "file" the latter; (*h*) knowledge need not make us more receptive: learning develops a "set" (*Einstellung*) or habit that makes if difficult for the subject to face new problems or situations; (*i*) we may be insufficiently trained in logic, scientific method, statistics, and experimental design—all of which help learn from experience; (*j*) we may need to appear in the right for social reasons, and may thus be driven to simulate that we have got the message, or that there was no message to be received.

Some philosophers, particularly Popper and his followers, believe that we learn only from our mistakes, i.e. by punishment. Experimental psychology has refuted this view, showing that most people manage to learn a lot and yet they disregard negative evidence and seldom seek it (Wason, 1960; Smedslund, 1963; Jenkins and Ward, 1965; Einhorn and Hogarth, 1978). Consider the following situation. A subject is instructed to take action A if a certain variable X lies below cutoff point X_c, and action B otherwise. Further, his decisions have outcomes of two kinds: success in case a certain variable Y falls below a certain cutoff point Y_c, and failure otherwise. The state space of Figure 1.11 can then be built. Experimental

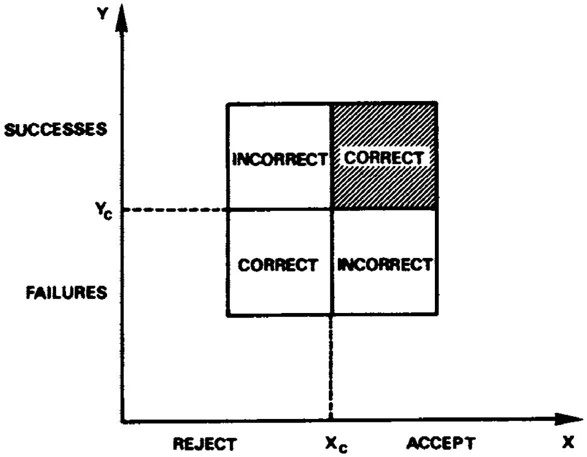


Fig. 1.11. A common fallacy. See text. Adapted from Einhorn and Hogarth (1978, p. 397).

studies (Einhorn and Hogarth, 1978) show that the only feedback most people care about is that coming from the pairs ⟨correct decision, success⟩, i.e. the points in the shaded cell. They generally ignore information in the three other cells, i.e. they tend to disregard the misses—as Bacon noted—and even the correct forecasts of failure (or non-occurrence). In other words, most people do *not* learn from their mistakes or even from their hits unless the latter bring rewards. Only a scientific training may instil the habit of seeking disconfirming information. In other words, ordinary experience is not self-correcting: it only gives us a chance to recognize error and learn to avoid it in future. Only scientific experience is consistently self-correcting.

If we knew more about learning we would not propose simple-minded epistemologies and would be able to learn in more efficient ways. As it is, we must make do with a handful of rules of thumb. One such rule is that *extremes of speed must be avoided in learning*: too slow a pace because it induces boredom, too swift a rhythm because it leaves no time for consolidation. We must regulate the inflow of information to keep it within moderate bounds. It must be quick enough to keep the neural systems alert to novelty: otherwise they become habituated or adapted, i.e. they do not respond. And the flow must not be so quick that psychons are dissolved no sooner they are formed, with the resulting confusion and poor recall. Low speed is best left to invertebrates, and high speed to computers: we had better choose medium speed and concentrate on selectivity, creativity, generality, and depth. Selectivity is becoming more and more important because we are subjected to an incessant bombardment with irrelevant information: we must learn to evaluate and discard or skim most of the information that comes our way if we wish ever to make an original contribution to human knowledge. We must learn, in short, to choose wisely what to learn: "what we all need at this point in human evolution is to learn what it takes to learn what we should learn—and learn it" (Aurelio Peccei).

A second rule for efficient learning is: *Avoid cramming huge piles of unrelated data* ("facts"), for they are difficult to comprehend and recall. We tend to organize the stream of incoming information into "chunks" that can be understood and recalled as units or systems (George A. Miller). Thus coherent stories can be understood and recalled more readily than isolated sentences, and the latter more easily than nonsense syllables. The keys to this are of course connectedness (systemicity) and significance (what signs "say" about what). If this finding of experimental psychology were better known, chemists, biologists, medical researchers, social scientists and others might lower their resistance against theorizing, for theories are systems of knowledge. But of course theories are not data packages and therefore they will never replace data, such as addresses, formulas of chemical compounds, or historical dates. Hence we must learn such items alongside theories—but learning the latter will ease the learning of the former, for theories can be used as filing systems where data can fit.

A third rule to be followed is this: *Don't miss the apprenticeship period and get hold of the best teachers you can.* All animals capable of learning pass through a period of apprenticeship in some skill or other—e.g. flying, singing, fighting, exploring the environment, earning a living and, in the case of primates, even copulating. One does not become a farmer or a master builder, a manager or a research chemist, just by reading books, auditing courses, or watching television programs: every craft is learned at the knee of some master. One reason is that know-how cannot be wholly transmitted except by example. Another is that the text-book and the research paper give condensed and tidy reports on a process that has been anything but brief and tidy. A third reason is the need for incessant feedback (correction and encouragement). Those who object to apprenticeship calling it paternalistic, coercive, or repressive, will never train any disciples except in the arts of barren revolt and misleading rhetoric.

However, stressing the need for apprenticeship should not be mistaken for praising brainwashing, or even for favoring protracted apprenticeships. It is true that humans are remarkably immature at birth, and that they have to learn to cope with a habitat made extraordinarily complex by their forebears. Still, it is questionable whether the current apprenticeship period is wise, the more so since it tends to increase. Doctorates in science are obtained in the West on the average at 29 years of age, i.e. they take 23 years of schooling not counting preschool. The reasons for this long and costly apprenticeship period seem to be to make sure that the candidate (*a*) has mastered the skills of the previous generation—which is unnecessary and impossible anyway; (*b*) will stick to the traditional ways (rather than try radical novelties), and (*c*) will stay out of the market place, thus protecting his teachers and himself. The apprenticeship should not be so short as to leave the student ill prepared to face professional problems by himself; but it should not be so long as to dry up every innovative and rebellious impulse. We should start earlier and be more demanding; but we should demand less curricular uniformity, cram less stray data, and allow more freedom of search and criticism.

Finally, are there limits to what an individual can learn? Is every one of us, as information processing psychology claims, a limited capacity information processor subject to Shannon's law concerning the limited capacity of every channel (Miller, 1967)? The limited capacity thesis applies to telephone lines, computers, and other artificial information systems, but it has not been proved for humans. There are indications that it does not apply to our brains: (*a*) unlike artificial information processors, our brains are plastic and, in particular, they have the property of self-organizability; (*b*) every time we learn something we become better prepared to learn further items: learning is an autocatalytic process, not the filling of prefabricated book shelves; (*c*) with practice we can learn to do two or more tasks at once—e.g. driving and chatting, typing and listening to music (cf. Neisser, 1980). Of course there are limits, some biological (e.g. poor genetic endowment, loss of neural plasticity, short life span) and others social (e.g. poor cultural environment, few opportunities). But these limits are not fixed once and for all: our brains and societies are to a large extent self-made. We can push the limits to what we can learn by building a society where everybody can freely exert the right to learn. We shall return to this theme in Ch. 13, Section 4.

3. Development and evolution

3.1. *Cognitive development*

Organisms develop from birth and biopopulations evolve over time. Hence there are three different though mutually complementary ways of studying biofunctions, in particular cognitive abilities: synchronically, developmentally, and historically. The synchronic approach attempts to learn what and how an animal learns; the developmental studies the growth and decline of cognitive abilities: how they are formed at an early age, what sequences they constitute, etc.; and the historical or evolutionary approach studies the origin and subsequent evolution of cognitive functions. Clearly, the synchronic study must initially precede all others, for it is impossible to discover the development or the evolution of X unless one has some idea of what X is. But once one has learned something about the development or evolution of X, one can form a better idea of what X is—if only because X has developed and evolved from some predecessor that shares some properties.

All the components of our body are genetically programmed, but they are fashioned, within the genic constraints, by the way we live and in particular the things we do. This holds for the muscles: compare those of an athlete and a sedentary professor, or a beef eater and a tortilla eater. And it holds for the brain, as the neurons of the plastic components of the brain can sprout dendrites and make synaptic connections all the time—or lose them with disuse or age. The brain, like the muscles, develops the way it is nourished and used. Thus our entire body is to a large extent self-constructed—particularly so the brain, which ultimately selects what food and what information it is to feed on. We are born with the ability to form new neural assemblies or psychons, but which ones will be formed depends on both internal and environmental circumstances. A severe protein deficiency in infancy may result in low intelligence, and repeated exposures to antisocial behavior may ensue in the same. The genome determines only the potential: which potentialities are actualized depends on the way the psychons develop, and this development is molded by both the natural and the social environments.

The brain of the neonate is immature: it weighs only one-fourth of the adult brain, and its cortex performs few specific functions. In particular its cortical neurons are small, have few and short dendrites, and are hardly interconnected. No wonder the newborn human is slow, clumsy, mindless, and therefore helpless. As the child grows up its neurons grow, sprout dendrites, and make synaptic contacts with neighbors. What began looking as a sparsely populated forest ends up by looking like a tangled jungle: see Figure 1.12. The maturation process continues throughout life unless senility sets in.

The maturation of the brain is a process composed of five main threads:

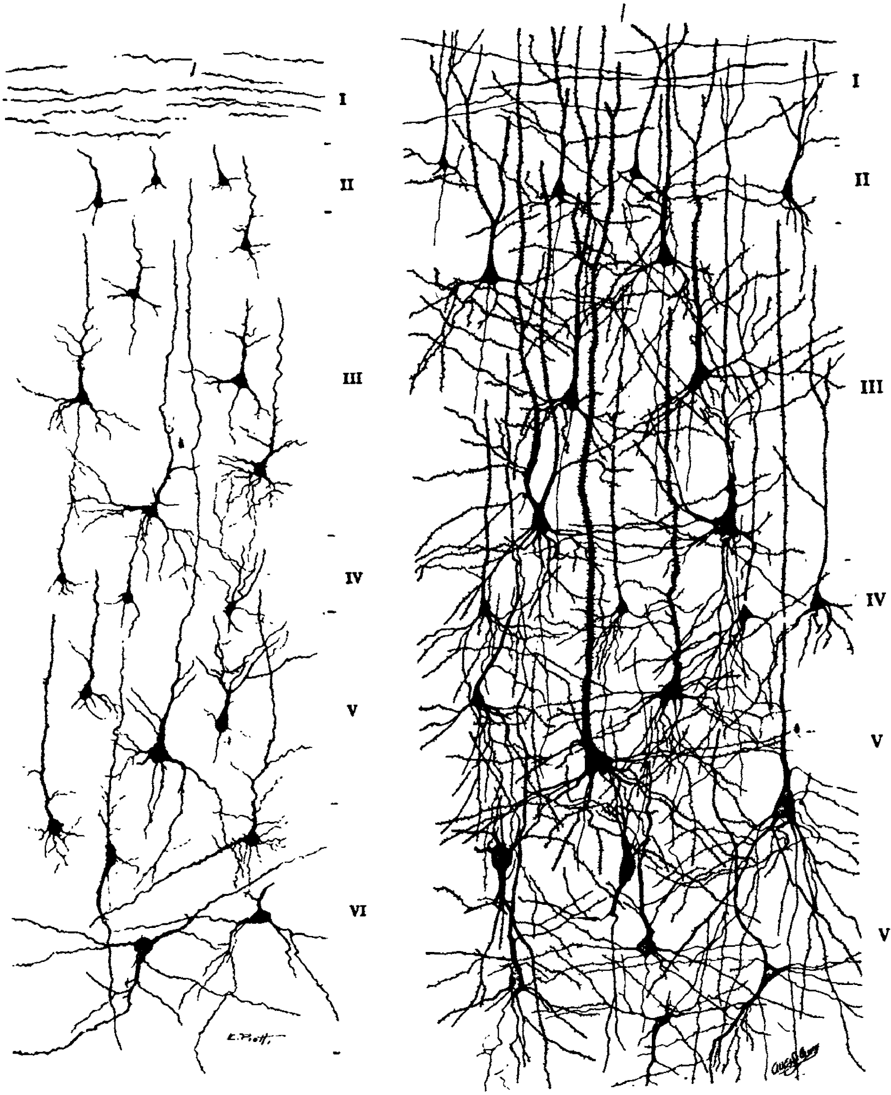

I
II
III
IV
V
VI
I
II
III
IV
V
VI

Fig. 1.12 Sections of the same region of the cerebral cortex of a newly born human (left) and of a 7 year old. From Conel (1939/67).

(*a*) growth of neurons; (*b*) growth of the myelin sheaths that wrap up and insulate the neuron axons; (*c*) change in the chemical composition of the neurons and the synaptic clefts; (*d*) proliferation of dendrites, and (*e*) multiplication of possible contacts with neighboring neurons, and thus formation of new psychons. Whereas processes (*a*) and (*b*) are completed in a few years, the others seem to go on for life. The maturation process is not homogeneous; in particular, the various cortical "areas" mature at different rates. Thus the motor "areas", which are partly prewired at birth, mature earliest; then come the sensory "areas", and in the first place the visual one. It is estimated that the cortex does not become fully functional until around 20 years of age. And it is known that the connectivity or interneuronal linkage remains changeable throughout life—barring senility. As the maturation process proceeds, new abilities emerge whereas others are lost. Thus it is likely that every one of Piaget's stages is the manifestation of the attainment of maturity (full functionality) of a large neuronal system.

The development of the nervous system can be somewhat accelerated by favorable environmental circumstances, and it can be distorted or even arrested altogether by unfavorable ones. For example, a kitten allowed to use a single eye during its first year of life will fail to develop binocular vision even after having been allowed to use both eyes (Wiesel and Hubel, 1965). A monkey shown only vertical lines in infancy will be unable to perceive horizontal bars later in life. A child who is not spoken to until age three may never learn to speak properly. In short, if the proper interneuron connections are not made in time, they are never made. In particular, certain cognitive abilities are not acquired or developed if the sensitive or critical period is missed. Practical consequence: Expose children and youngsters to rich and variable environments, and give them a chance to learn skills as early as possible. *Early training* is the ticket.

Developmental neuroscience and psychology, though still in their infancy, are in a position to solve the old nature/nurture (or nativism/empiricism) controversy, i.e. the question whether our cognitive abilities are inherited or learned. (The current dispute over the inheritability of intelligence is of course part of that controversy.) However, before rushing to take sides, or to propose a third alternative, we should clarify the rather vague notion of innateness.

According to contemporary biology "innate" subsumes two different concepts: "inborn" and "inherited though not present at birth". An inborn trait is one possessed at birth, such as the capacity to breath, move, cry, look, or listen to spoken words in the case of human infants. Most inborn traits are inherited but some are outcomes of genic changes (e.g. mutations) or of original embryogenic processes. An inherited trait is one handed down by the animal's ancestors through the DNA molecules contained in the germ cells. Not all inherited traits are inborn; many emerge (or are "expressed") during development until the adult stage is reached. And, whether present at birth (inborn) or expressed later on (inherited), every such trait is ultimately controlled by the hereditary equipment (genome). In particular, an innate (inborn or inherited) cognitive function is one performed by a genetically programmed system. Thus "Animal *A* possesses innate capacity *B*" is spelled out "Animal *A* possesses neural system *C* which, when activated, discharges function *B*".

Now, like any other material system, from atom to society, a neural system has properties of two kinds: actual and potential. I.e. there are some "things" it does and others that it may do. Thus an atom is always moving relative to some thing or other (actual property), and it has the possibility of being excited by incoming radiation of certain wavelengths or, if already in an excited state, of decaying to some lower energy level. Likewise a neural system has, in addition to some permanent properties, others it can acquire, such as learning to perform certain functions. Indeed, whereas some neural systems are prewired (or develop into rigidly wired systems), others are plastic, i.e. they are formed, strengthened, or weakened in the course of life. (Recall Section 1.1.) Thus an infant starts learning the moment it is born: this shows that the ability to learn certain skills is inborn. On the other hand, in order to learn other skills the animal must wait for its nervous system to mature, and for the environment to offer the opportunity. However, this opportunity may never arrive, particularly in conditions of extreme biological and cultural deprivation. Suffice it to recall that children raised with insufficient food, parental care, perceptual and mental stimulation, and interaction with their peers, do not develop a normal brain: they can seldom if ever become intelligent.

In the light of the preceding we realize that nativism and empiricism, the psychological trends dealing with the nature/nurture controversy, are mutually complementary rather than exclusive (Dodwell, 1975)—and yet they are not exhaustive because neither acknowledges creativity. The reason both are needed is that, in the light of psychobiology, each is concerned with a certain kind of neural system. Nativism is (tacitly or explicitly) concerned with wired-in or rigid connections, which are typically innate (either inborn or formed later under the control of the genome). On the other hand empiricism is (tacitly or explicitly) concerned with plastic or variable connections, which are typically acquired (learned).

Since the brain of a higher vertebrate (mammal or bird) contains systems of both kinds, we need empiricism as well as nativism—or rather a merger of the two enriched with the hypothesis of creativity, or the capacity of learning beyond experience. Only such a doctrine could account for the fact that we are born with a central nervous system that is partly rigid (prewired or programmed) and partly plastic (acquired or learned). But this is not all: although our plastic neural systems are partly self-constructed (by learning) they also have inherited capacities. Just as no two animals have exactly the same set of proteins—hence exactly the same chemical reactions—so no two animals are likely to have exactly the same neural systems and therefore exactly the same cognitive abilities. Such inherited neural differences, together with differences in environment and formal education, explain why some children are better than others at games or at languages, at mathematics or at drawing.

(A recent review of 111 studies published in the world literature, and covering 113,942 pairings, confirms the so-called polygenic inheritance of intelligence hypothesis, according to which the higher the fraction of genes two family members have in common, the higher the average correlation between their IQ's (Bouchard and McGue, 1981). However, the data exhibit also strong environmental influences. Thus whereas the average correlation between monozygotic ("identical") twins reared together is 0.86, that of monozygotic twins reared apart is 0.72; the corresponding figures for siblings reared together and apart are 0.47 and 0.24 respectively. Surprisingly, the figures for nonbiological sibling pairs are: 0.29 for adopted–natural pairs, and 0.34 for adopted–adopted pairs. Also, there is no evidence for the role of sex on general cognitive development. Finally, according to another study (Loehlin and Nichols, 1976), "identical" twins correlate only about 0.20 higher than fraternal twins in personality traits.)

In sum, the nature/nurture disjunction is not exclusive: genes determine only potentialities that are actualized or thwarted by the environment. What about innate ideas: are there any? Was Socrates right in holding that there is only remembrance, not acquisition of new knowledge? Before discussing this matter we should clarify the meaning of "idea", the more so since in the philosophical literature instincts and perceptions have sometimes been lumped together with images and abstract ideas. For the sake of precision we shall restrict ideas to concepts (e.g. "wet") and propositions (e.g. "The ground is wet."). The problem can now be restated as follows:

Are any concepts or propositions inborn or inherited? Innatism replies in the affirmative, empiricism in the negative. This 17th century debate has been reactivated in recent times in connection with the psycholinguistic problem of language acquisition—of which more in Ch. 3, Section 3.2.

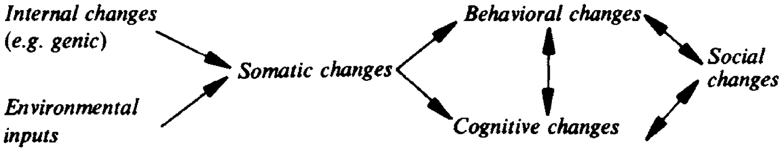
The controversy about innate ideas is bound to remain unresolved as long as it is conducted on the superficial level of traditional mentalistic psychology. Neuropsychology (or psychobiology) is in a position to make a decisive contribution by suggesting that we study the language cortex and its function from birth to school age. We still know very little about this subject: psycholinguists seem to have been busier speculating and polemicizing than studying the child's brain and social environment. However, we do know this much: (*a*) ideation is an activity of fairly large neuron assemblies (psychons), not of germ cells, let alone of their DNA content, which is the only thing that gets transmitted from one generation to the next; (*b*) the brain of the newborn is so poorly organized (recall Figure 1.12) that it cannot even coordinate its movements, so that it is unlikely that it can form any concepts or propositions; (*c*) development is not merely the unpacking or "expression" of genic destiny: it is a creative process punctuated by the emergence of new functions and the exercise of some such functions in the acquisition of new knowledge; (*d*) children who grow up in culturally deprived environments have poor sensory-motor skills and a very poor inner life: although ideation is a brain function, it is not discharged in a social vacuum. In sum, we do not inherit knowledge but only the ability to acquire it in appropriate biological and social conditions. Which is just as well, for otherwise we would be lacking in curiosity, we would get stuck with ineradicable errors along with genetic defects, and the advancement of knowledge would have ceased long ago.

None of this entails that the newborn's brain is a clean slate ready to accept any inscriptions. On the contrary, it is partly a plastic brain with definite potentialities: it inherits not only certain rigid circuits (in charge of the basic household functions), but also the capacity to form new ones, as well as of dismantling some of them. Moreover we can form some such neural systems at will, as when we decide to learn a new motor skill, a new language, or a new branch of knowledge. In sum, we are unlikely to be born with any ideas, but we are born with the possibility of learning and even creating new ideas. (Maudsley (1876, p. 278) wrote that the notion of an innate idea is as absurd as that of innate pregnancy.) Shorter: although we are born ignorant we possess the neural apparatus to learn and, given the chance, we do learn quickly almost anything. Of course there are individual differences traceable to genic differences: some people learn music, philosophy, or engineering, more easily than others. But this is beside the point: the point is that all such skills must be learned.

We share the instinct to learn with numerous animal species. However, we have a definite advantage over them, namely that of being born into a culture built by our ancestors. This inheritance spares us the need to make a fresh start in every single respect no sooner we learn to walk. (One's school record does not recapitulate the history of mankind.) But every cultural inheritance is ambivalent: it assuages one man's thirst for knowledge and incites another's exploration of new territory. Besides, whereas some societies encourage learning, others inhibit it, or at least they discourage inquiring into certain subjects. This being so, we will not learn much about learning if we restrict our inquiry to individual brains: we must study also the social matrix of the learner. More on this in Ch. 3, Section 2.

3.2 *Cognitive Evolution*

The evolution of cognitive abilities is an aspect of biological evolution and social history. All evolution proceeds by internal (in particular genic) variation and environmental (natural or social) pressure. Both factors may be pictured as inputs on the organism. In the case of animals the outputs of such externally induced or modulated changes are modifications of various kinds. Two such kinds of change are of paramount importance for the student of cognition: behavior (in the narrow sense of outcome of muscular activity) and cognition. These outputs are interdependent: new motor capabilities open up new possibilities of perception and ideation, and conversely. And both outcomes, the behavioral and the cognitive, are molded by society and in turn contribute to molding the latter.



The study of the evolution of cognitive abilities is of course one of the tasks of evolutionary psychology. Unfortunately the latter is still underdeveloped to the point of being almost exclusively speculative. There are two main factors of underdevelopment. One is the scarcity of data, almost all of which consist in circumstantial evidence. The other obstacle to the progress of evolutionary psychology is psychoneural dualism: by detaching mind (or "software") from brain (or "hardware"), one forgets that what behaves or ideates is always some animal or other, and that animals happen to be subject to evolution. Nevertheless palaeoanthropologists and primatologists are beginning to learn a good deal about the evolution of the cognitive abilities of our ancestors over a time period of nearly three million years. (See Jerison, 1973; Masterton *et al.*, 1976a, b.) Such problems and findings should be of interest to all biologically oriented psychologists. By the same token they are a matter of indifference to mentalist psychologists, in particular those of the information processing variety, who claim to study either immaterial (disembodied) minds or computer programs, neither of which are subject to organic evolution.

Evolutionary psychology, then, is an underdeveloped branch of evolutionary biology. However, we cannot treat the evolution of the higher vertebrates in the same manner as that of the lower vertebrates, let alone that of invertebrates. Indeed the search for knowledge and possession of it modifies behavior, and behavior is not only an outcome of evolution but also a motor of it (Piaget, 1976b). In other words, by generating and using new knowledge the higher vertebrates can improve their survival strategies and thus modify the direction and pace of their own evolution. This being so, it is not true that for them, as for the lower animals, all variation is blind, and selection the only source of fitness—as assumed, e.g. by Campbell (1977). To some extent mammals and birds can control variations in themselves as well as in their environment, thereby improving their own fitness. Still, this does not entail that higher vertebrate (and particularly human) evolution is Lamarckian rather than Darwinian (as claimed by Piaget (1976b)). DNA, which is what we inherit, does not learn from experience. (For further criticisms see Bunge (1981a).)

A number of thinkers, notably Spencer, Helmholtz, Peirce, Mach, and Popper, have drawn a parallel between the history of ideas and biological evolution: in both cases there would be problems, trials and errors, and only the useful or adaptive novelties would "survive". This analogy passes sometimes for evolutionary epistemology, but it is not. To begin with it is a metaphor, not a theory. Secondly, it is a superficial analogy for, unlike organisms, ideas are neither alive nor self-existent, hence they do not evolve by themselves. New knowledge is not gained by random mutation, and scientific and technological knowledge is seldom gained by blind trial and error. Most new knowledge is obtained by inquiry, and there is no such search unless the animal is motivated (by need or by curiosity) and unless it knows some search strategies, however primitive.

A third objection to what passes for evolutionary epistemology among philosophers is that it is false that the evolution of knowledge consists simply in the elimination of "unfit" hypotheses, i.e. conjectures that fail to pass empirical tests. There is more to knowledge than hypotheses: there are also data and problems, viewpoints and methods, rules and instruments, designs and plans, etc. Fourthly, some patently false superstitions minted thousands of years ago—e.g. the doctrine that ideas dwell in a world of their own—are still going strong, whereas a number of scientific truths and useful technological innovations lie buried in the specialized literature, where they are likely to remain unless independently rediscovered or reinvented. Fifthly, truth does not always prevail once discovered or invented. It is sometimes suppressed for being incompatible with received dogma or vested interest, and at other times it is forgotten for lack of interest.

In short, most of what passes for evolutionary epistemology among philosophers is as removed from biology as it is from the history of knowledge. (For further criticisms see Currie (1978).) Genuine evolutionary epistemology takes organic evolution seriously, deals with the evolution of cognitive abilities as an aspect of brain evolution, and takes the social matrix into account. (For a sketch see Vollmer (1975).) True, evolutionary epistemology is so far little more than a research plan, so the philosopher cannot learn much from it. However, he should adopt a genuine evolutionary viewpoint, which involves dealing with concrete organisms immersed in concrete environments, rather than with disembodied minds (or computer programs) hanging in a natural and social vacuum.

4. CONCLUDING REMARKS

It is legitimate to *distinguish* knowledge from the cognitive process. But it is impossible to *separate* them or to try to understand the one without the other—as impossible as to understand ashes without fire. Philosophy may well leave the detailed study of cognition to the cognitive sciences, but it cannot ignore them, and it must make such studies the firm (though changing) root of epistemology. It can disregard them only at the risk of producing obsolete and barren armchair psychology.

There are several approaches to cognition. One is to treat it as a particular kind of behavior, and to describe the latter in terms of stimuli and responses. We reject this, the behaviorist approach, because it ignores that which does the cognition, namely the nervous system, and it does not even pose the problem of accounting for intelligent behavior. A second approach consists in begging the question: in taking ideas for granted and trying to account everything else in terms of them. We reject this, the mentalist approach, for being superficial and nonscientific: it fails to explain the nature of ideas, it assumes that they are self-existent, and it ignores the brain, thus isolating psychology and epistemology from all the other sciences. A third approach to cognition is the information processing one, i.e. cognitivism, which does not distinguish between brains and computers, is interested in the processing of information but not in its creation, and identifies body with hardware and mind with software. We reject this approach as well for failing to explain how information is generated and how knowledge guides (or misguides) behavior; it ignores the peculiarities of the nervous system (which it views as one more information processing system), as well as development and evolution, and it presupposes and therefore encourages psychophysical dualism.

The fourth approach to cognition is the *biosociological* one, the basic thesis of which is that every cognitive activity is a neural process, and one that interacts vigorously with other processes in the animal as well as with its natural and social environment. We adopt this approach for the following reasons (Bunge, 1980a, 1981a). First, the biosociological approach to cognition is in harmony with the ontology (or world view) of science, which is one of lawfully changing things rather than one of immaterial entities or disembodied processes. Second, it jibes with the life sciences, in particular (*a*) developmental and evolutionary psychology, which teach us that organisms change considerably in the course of their lives, and biopopulations may change to the point of speciation or extinction, and (*b*) neuroscience, which among other things can explain how the nervous system perceives, imagines, feels, thinks, etc.—or fails to discharge satisfactorily any such functions. Third, the biosociological approach to cognition harmonizes with the social sciences, in particular the sociology of knowledge, which shows that society molds the way we see the world and ourselves, and studies the ways in which society encourages or discourages inquiry, and benefits or suffers from its results.

CHAPTER 2

KNOWLEDGE

We may distinguish knowledge from the process of its acquisition: we may think of knowledge as a product or outcome of cognitive operations such as perception and inference. But of course there is no outcome detachable from process: the former is just the final phase of a sequence of events. Nor is knowledge the invariable outcome of a cognitive process: sometimes inquiry ends up in ignorance, at best in the resolve to engage in further inquiry. In any event knowledge, when acquired, is a collection of learned items: it is a collection of brain processes or a disposition to replay them. Yet although we cannot detach the outcome (knowledge) from the corresponding process (cognition), we may distinguish them.

Moreover for some purposes we may *feign* that cognitive processes have a "content" that can be communicated to other brains or externalized as artifacts such as inscriptions and tapes. To be sure actually there is no such content and, a fortiori, no such transfer. Acquiring knowledge is learning something, i.e. going through a certain brain process, hence not the same as acquiring a book or some other commodity. Likewise exchanging information is not like trading things but is interacting with another animal in such a way that each party elicits certain learning processes in the other's brain.

Yet from a methodological point of view it is convenient to feign that cognitive processes do have a transferable content, so that we may think of the latter separately from the former. This convenient fiction amounts to supposing that, for certain purposes—such as checking validity or usefulness—it does not matter what or who went through the cognitive process in question. It could have been Archimedes in Syracuse two millennia ago, or Einstein in Bern at the beginning of our century. This distinction between cognition and its content is useful in epistemology and there is no harm in it as long as it be understood as a fiction, not as an ontological thesis on the autonomy of ideas (Bunge, 1981a). After all, mathematics too deals with fictions—though not wild or idle ones—and this does not detract from its usefulness. On the contrary, the distinction guarantees impersonality and facilitates universality.

We shall also distinguish knowledge from belief, but this distinction is not a mere methodological gimmick. That knowledge is not belief should be clear from reflection on cases such as "*A* knows *B* but does not believe *B*", and "*A* believes *B* although he does not really know *B*". Yet there is a large body of philosophical literature that defines knowledge as a kind of belief; we shall pay hardly any attention to it.

Finally we shall sketch the features of the deliberate and systematic acquisition of knowledge through inquiry, leaving its detailed study to later chapters. And we shall characterize a field of inquiry in terms of a material framework and a conceptual one.

1. FROM COGNITION TO KNOWLEDGE

1.1. *Knowledge Acquisition and Use*

We have assumed that every cognitive activity is the activity of some plastic neural system or psychon (Ch. 1, Section 2). We must now distinguish between acquiring (learning) a piece of knowledge and using it (e.g. recalling it or putting it into practice). From our neuropsychological point of view this distinction amounts to distinguishing the *original formation* of new psychons (or psychons of new types) from the *activation* of existing psychons (or the formation of psychons of a type that already had representatives in the individual's brain). In other words, we distinguish between the assembly of new neural systems (or circuit rewiring, in the old metaphor) and the use of existing ones. Presumably the lower animals can do only the latter: they do not self-assemble or self-organize new neural systems, except perhaps through malfunction. But we do it as long as we learn new items.

Example 1: An animal is in the process of learning to pair stimuli of kind E to responses of kind R, each accompanied by internal states belonging to the brain state space S. (For the notion of a state space see Ch. 1, Section 1.2.) That is, the brain of the animal is establishing a new psychon "embodying" (represented by) a state function of type $F{:}E \times S \rightarrow R \times S$. Note that this is a very strong assumption, for it implies that a finite number of stimuli can elicit the emergence of a property represented by a function F from $E \times S$, which may contain infinitely many elements, to $R \times S$, which may also be composed of infinitely many members. Clearly, this conjecture goes against the grain of inductivism, which demands that we learn item by item and synthesize only at the end. (More in Ch. 5, Section 2.1.) And the hypothesis is equally alien to innatism: in fact it is an alternative to both rationalism and empiricism.

Example 2: The animal has already acquired the function F in the former example, and so it has acquired the *disposition* to make responses of type R every time it receives stimuli of type E. It is now subjected to a stimulus of this kind (though possibly different from any of the individual stimuli that induced its learning F), and it generates both the suitable response and the accompanying inner state. It is making use of a bit of knowledge acquired earlier, namely the one summarized by the function F.

To give some precision to the preceding ideas we introduce a certain notion of functional equivalence. We say that the linear functions $y = x$ and $y = ax + b$ are equivalent, or belong to the equivalence class of linear functions. Likewise the trigonometric functions $y = \sin x$ and $y = \sin (ax + b)$ will be said to belong to the equivalence class of sine functions. In general, we say that two functions differing only by a linear transformation of their independent variable(s) are *equivalent*. And we write $[F]$ for the *equivalence class of F*, i.e. the set of all functions equivalent to F. We also say that, if the state functions of two concrete (material) systems are equivalent, then the systems themselves and their activities (i.e. the processes occurring in them) are *functionally equivalent* no matter how much they may differ in details.

We are now ready for our definitions of knowledge acquisition and use. Let b be an animal endowed with a plastic neural supersystem P, and consider a time interval τ. We stipulate that

(i) *b acquires knowledge of kind* $[F]$ over period τ if and only if a new psychon, characterized by a state function F (a member of $[F]$), is formed in P in the course of τ;

(ii) *b makes use of knowledge of kind* $[F]$ over period τ if and only if the P of b contains at least one psychon, active during τ, characterized by a state function of kind $[F]$ at the beginning of τ;

(iii) the *total knowledge* of b at time t is the union of all the bits of knowledge, of all types, that b has acquired up until t. We call this set $K(b,t)$.

Note that, contrary to the behaviorist tradition, but in consonance with cognitivism, we have equated learning with the acquisition of knowledge (see, e.g. Dickinson, 1980). However, in contrast to mentalistic cognitivism, we equate in turn the acquisition of knowledge with a partial restructuring of the nervous system rather than with the formation of a new immaterial mental configuration. Our hypothesis is more precise and therefore better testable than the mentalistic one, which contains the essentially vague notion of a mental configuration. Moreover our hypothesis has recently received a sensational experimental confirmation from the study of bird song. In fact, it has been found that in the spring, when male canaries learn new song repertoires, their song control neural nuclei nearly double in volume relative to the fall, when they stop singing (Nottebohm, 1981). This seems to indicate that learning is (identical with) the growth of new dendritic segments and the consequent formation of new synapses, whereas forgetting is the shedding of synapses.

Note also that our definition allows for considerable variability among individuals, and even in the course of the lifetime of any given individual. Thus even if two individuals learn "the same thing" (e.g. to shake hands, or to read graphs), there may be considerable differences in the details of their learning processes. Consequently there may be differences in the outcomes or performances, e.g. in the way they shake hands or read graphs. Second and more important, learning is not done by the mythical hollow organism of behaviorism or by the mythical disembodied mind of the mentalist: learning consists in some nervous system acquiring new properties. (Other organs, such as the heart and the kidney, do not have this capacity: they have fixed state functions.) Third, there is no instantaneous learning, hence no instant knowledge: getting to know anything takes time, however short.

Our definitions have a number of consequences, some of which are irrelevant to traditional epistemology whereas others are incompatible with it. Here are some of them:

(i) An animal lacking any plastic neural systems does not know anything. (Thus a worm has no more knowledge than a stone. When it learns to withdraw from a noxious stimulus, all it does is to inhibit or block certain neural pathways: it does not create any new ones.)

(ii) An animal whose plastic neural system fails to contain a subsystem characterizable by a state function of kind $[F]$, and active during τ, has no knowledge of kind $[F]$ during τ. (Inactivity may be due to sleep or coma, or it may be caused by drugs.)

(iii) All cognitive deficits and abnormalities are dysfunctions, temporary or permanent, of the corresponding neural systems. (Such dysfunctions may occur at the molecular level, as is the case with excesses or deficiencies of neurotransmitters, or at the cellular level, as is the case with "wrong" synaptic connections.)

Acquiring knowledge and using it are cognitive activities, but not every cognitive activity is successful: we may conduct an investigation without learning anything. In other words, the collection of learning activities is properly included in that of cognitive activities. Consider the collection of all the learning activities of an animal b up until time t, i.e. the total knowledge $K(b, t)$ of b until t. We obtain the knowledge of the species to which b belongs by forming the union of all the individuals' knowledge of the species. Thus what we call *human knowledge* at a given moment is the union of all the bits of knowledge—correct or incorrect—acquired by members of the human species up until that moment, i.e. $K(H,t) = \bigcup_{b\in H} K(b,t)$. This is a large collection, for it includes every truth, half truth, and error that every single human has ever learned. We risk nothing by postulating that nobody will ever learn all of human knowledge. Which is just as well, for much of it is not worth acquiring anyway.

For most of prehistory knowledge existed only in individual brains. With the invention of drawing, painting, sculpting, and particularly writing, knowledge could be encoded and externalized in *cultural artifacts* that could circulate throughout the community. This facilitated the storing, sharing, and enriching of knowledge. But it also fostered the myth of the independent "content" of knowledge—independent, that is, from the inquiring subject.

It is easy to see how this myth can be generated and maintained. When somebody finishes a paper or a drawing, these pieces of matter can be detached and seen by somebody else: even their creator can stand back and contemplate them as if they were self-existing, while in fact their "content" depends on their being perceived and understood by some brain. This creates the illusion that we are in the presence of three separate items: the neural (and motor) process resulting in the writing or drawing, the cultural artifact, and the knowledge or feeling encoded in the latter. The next step is to collect all such bits of knowledge detached from brains—i.e. all the problems and data, theories and plans "in themselves"—and endow such a collection with a life of its own. The final step is to give a name to such a collection of items allegedly hovering above brains and society—e.g. the "realm of ideas", the "objective spirit", or "world 3". (See, e.g. Plato, Hegel, Bolzano, or the later Popper.) In this way the illusion is created that such "worlds" of knowledge and feeling persist or subsist once they have been created by concrete individuals in specific social circumstances, and moreover that they can interact with living beings. Actually those collections do not constitute worlds (systems), for they are quite heterogeneous, particularly if cultural artifacts are counted among their inmates. Worse, there can be no empirical evidence for the hypothesis that such ideal "worlds" lead an existence separate from living brains. (Cf. Bunge, 1981a.)

To return to learning. The statement that somebody has learned a bit of knowledge is ambiguous. It may signify that she has created some new knowledge, or that she has grasped a bit of "existing" knowledge—existing, that is, as a process in someone else's brain or encoded in some cultural artifact such as a book or a film. The difference is that between production and consumption, and it is both a psychological and a social difference. It is psychological because producing new knowledge is generating psychons of a type that had not existed before in the evolution of the species, whereas to learn a bit of existing knowledge is to generate a psychon similar to one that had already been activated in someone else's brain. The difference is also social, for 'original knowledge' means "knowledge not available to anyone else before".

A bit of knowledge, whether original or unoriginal, may be thorough or partial, and it may be of high or low grade. Thus we may have partial or full knowledge of a gossip or of a theorem. So learning, whether afresh or second hand, comes in degrees. (Moreover we shall postulate that only constructs may be fully known. But this must await Section 5.) Unfortunately we still do not have a measure of degrees of knowledge.

As for quality, recall the (unclear) difference between knowledge and wisdom. The wise person is eager to learn discriminately. Knowing (having learned) much is not the same as being wise. A person may be a mine of information, yet quite incapable of creating new knowledge or correctly evaluating knowledge produced by others. Unlike erudition, wisdom is quality knowledge: it is knowing important things and knowing how to find out more of the same. Beware of those who claim that wisdom is disjoint from, and superior to, knowledge: they are likely to be superstition mongers. When we hear T. S. Eliot complaining "Where is the Wisdom we have lost in knowledge?", we should reply: "And where is the knowledge you parade as wisdom?"

Our view of knowledge differs radically from that of the ordinary language philosophers who hold that to know *X* is to know how to talk about *X*. Such knowledge would be "embodied" in certain speech rules rather than in bodies of data or in theories. Thus we are told that we may use the word 'quark' just in case there is a rule on how to use it. The stipulation has the advantage that it enables anyone with a command of ordinary language to pass for an expert on anything that can be talked about in the language.

Our view differs also from the empiricist doctrine according to which every learning process is gradual and proceeds by trial and error. We have made no such assumption, although we did assume that every learning process takes time. For all we know a number of higher vertebrates can learn certain "things" at the first go. Besides, all learning is driven by motivation and expectation. Also, not all trials are tests: errors are not easily eliminated even after ending up repeatedly in disaster. Just think of war and prejudice. The weight of prejudice is such, that learning new ideas often requires getting rid of false preconceptions (Bacon's *idola*). Thus one of the main obstacles to the learning of mechanics is the old superstition that a moving body eventually comes to rest for exhaustion of a driving force, fuel, or impetus (McCloskey *et al.*, 1980). Likewise the science student who wishes to learn some philosophy must begin by criticizing two views he learned in his science textbooks, namely that all concepts must be defined operationally, and that theories are data summaries.

Finally, our view is at variance with the conception of "knowledge as a matter of conversation and of social practice, rather than as an attempt to mirror nature" (Rorty, 1979, p. 171). To be sure, communication disseminates knowledge and social practice checks it: there is no cognition in a social vacuum. Still, cognition is a brain process not a social one, and we must have some information before we can trade it in conversation. Likewise there is no respiration in a physical vacuum, but this does not prove that respiration is an atmospheric or an ecological process. Besides, what would be the point of inquiry if it did not result in a representation of nature (and society)? The view of knowledge as a matter of conversation and social practice may fit the sophist but is a caricature of the investigator, be he a scientist, a technologist, or a humanist.

1.2. *Knowledge Evolution*

Our qualitative representation of knowledge in the preceding section allows us to speak of the *state of knowledge* attained by an animal at a given time, namely by identifying such state with the total knowledge acquired and retained by the animal up until that time. We denoted such knowledge by $K(b, t)$ (Section 1.1).

This representation allows one to compare any two cognitive states, be it of the same animal or of different animals at different times. Such comparison may yield, in principle, either of the following results. One of the animals knows *more* than the other (or the same animal at a different time), the two *share* some knowledge though not all, and the two do not share any knowledge. The first case might be that of the lower animals capable of learning something; the second, that of all primates; and the third pigeonhole seems to be empty. We assume then that *any two humans share some knowledge* and that *every one of them knows something that the other does not.* Without some common knowledge we could not communicate with one another; without some idiosyncratic knowledge there would be no point in communicating.

(The comparison of cognitive states can be synchronic or diachronic. In particular it allows us to estimate the gain or loss of knowledge in the course of time. Calling $K(b, 1)$ the cognitive state of animal b at time 1, and $K(b, 2)$ that at time 2, we may say that the *knowledge gain* between times 1 and 2 is $\gamma_b(1,2) = K(b, 1) - K(b, 2) = K(b, 1) \cap \overline{K(b, 2)}$. The gain is null, i.e. $\gamma_b(1,2) = \varnothing$, if the animal has learned nothing during that period. What holds for individuals holds, *mutatis mutandis*, for groups, upon introducing the definition of group knowledge as the sum total (union) of the knowledge of its members. In obvious symbols, $\gamma_G(1, 2) = K(G, 1) - K(G, 2)$, where $K(G, t) = \bigcup_{b \in G} K(b, t)$.)

The same formal tools allow us to compare (in principle) the knowledge states of different communities, or even of mankind at different times. If we do so we may find that there are always gains and losses. Thus we know many "things" that our ancestors ignored, but we also ignore much that made the existence of our forebears possible (and of others that made it unnecessarily harsh). In any case individual knowledge develops and human knowledge evolves: the former normally grows, the latter grows over certain periods but not over others. (Caution: knowledge can neither develop nor evolve, for it is not a concrete thing. Only individuals develop and communities evolve. However, the expressions 'development of knowledge' and 'evolution of knowledge' are permissible as long as they are understood to be elliptical.)

There are several views on the evolution of human knowledge. They can be grouped into two classes: those which hold, and those which deny, that the evolution of human knowledge follows definite patterns. Let us take a quick look at two well-known members of the former class. The best known of them is dialectics, according to which every construct carries within itself the seeds of its own negation, so eventually turns automatically into its opposite. This opposite would at the same time cancel and preserve the previous stage, and would eventually be subject to the same fate, i.e. negation. This second negation, far from being identical to the first construct (or thesis), would synthesize the first and the second: it would contain whatever was valuable in them, so it would occupy a higher rung in the ladder of progress. (Do not ask for more precision or for the mechanism whereby one item gets transmogrified into its opposite: dialecticians are neither clear nor thorough. The doctrine was proposed several millennia ago and has never gone beyond the verbal and metaphorical stage.)

The dialectical schema of the evolution of ideas is attractive at first blush: it is simple, it promises continuous progress, and it can be exemplified provided one is not too exacting concerning the notion of dialectical negation. However, the doctrine cannot be taken literally, for the only clear case of negation of an idea, namely logical negation, is not dialectical. In fact a proposition and its negation do not join to form a higher synthesis but only either a worthless contradiction or a harmless tautology. Also, it is an idealistic thesis, for it concerns the development of ideas in themselves (i.e. nonembodied ideas) rather than that of brains or communities. Thirdly, dialectics is confirmed by some examples but refuted by others. Think of the many futile controversies—theological, political and philosophical—that have occupied so many good brains for so many millennia without ever yielding any new, let along superior, knowledge. Or think of the way the great scientific theories have been created. What was the dialectical movement (of ideas) that ensued in Archimedes' statics, or in Newton's dynamics, or in Darwin's theory of evolution, or in Maxwell's electrodynamics, or in quantum mechanics? No dialectician has taken the trouble of telling us. In conclusion, there may be examples of dialectical process in the evolution of knowledge but there are also counter-examples, so there can be no general patterns or laws of cognitive dialectics. And, since there are no such laws, the principles of dialectics fail to explain or predict the evolution of knowledge. (More on dialectics in Bunge (1981a).)

The same holds for Popper's generalization of dialectics, i.e. what he calls the "method" of trial and error, summarized in the sequence: problem-theory-criticism (conceptual or empirical)—new better theory (Popper, 1972). For one thing this conception too is idealistic, for it deals with ideas in themselves in isolation from brains and societies. For another, not every theory, even if accepted, need be better than the theory it replaces: there are examples of retrogression or involution in the history of ideas. Thirdly, Popper's schema, which is a sketch of the usual conception of the scientific method, is normative not descriptive: it prescribes the way to go about investigating anything, but does not describe the actual cognitive evolution of mankind, or even of modern science. Hence it cannot be regarded as a generalization of the "laws" of dialectics. So much for two popular views on the evolution of knowledge.

It is doubtful that anybody has found *laws* of the evolution of knowledge. (See however Wilder (1981) for a set of alleged laws of the evolution of mathematics.) Consequently we do not even know whether there *are* any such laws. But we do know that, because ideas are not self-moving, we will not find any laws by examining ideas in themselves: we must investigate instead cognitive systems, i.e. learning individuals and communities, rather than the nonembodied "product" of their inquiry. This shift from ideas to their creators displaces the entire problem: it has now become a subproblem of the wider but more tractable problem of the laws of history. If there are such laws then the pattern of the evolution of knowledge is just an aspect of such laws. But, again, thus far we do not know whether there are laws of history. And in any case we must postpone this intriguing question to Vol. 6.

The result so far seems disappointing: we have succeeded only in proposing a qualitative representation of the state of knowledge and of cognitive change, and have expressed skepticism concerning the very existence of laws of such change. There are alternative approaches, two of which are quantitative. Let us see what they are and how they fare. The simplest of them is just to count knowledge items. This is not even a rough measure of knowledge, for the following reasons. One is that a mnemonist can memorize a huge amount of trivial information, whereas a foundational worker may be able to learn only a few but powerful items. Another, even worse difficulty, is the question of what is to count as a unit of knowledge: a proposition, a list of propositions (such as the one in a telephone directory), or a theory containing infinitely many propositions? In short, sheer numbers will not do.

A second and far more popular proposal is that of the subjectivist or Bayesian (or personalist) school of probability. It claims that (*a*) knowledge is a kind of belief, (*b*) belief comes in different strengths, (*c*) such strengths are measured by probabilities, and (*d*) probabilities are nothing else but degrees or strengths of beliefs, and so changes in probability values represent always changes in belief states. The centerpiece of the Bayesian school is the subjective interpretation of Bayes' theorem, which gives the probability of a hypothesis in the light of an item of empirical evidence, in terms of the prior (a priori) probability of the hypothesis together with that of the evidence and the likelihood of the latter given the hypothesis.

There are several objections to the theses that probabilities represent (only) states of knowledge or of ignorance, and the Bayes' theorem (or its information-theoretic rewrite) represents the learning process and even the evolution of human knowledge. Firstly, it leaves out problems and plans, methods and evaluations—in sum, whatever is neither hypothesis nor datum. Secondly, probabilities can be assigned to propositions only by decree: there are no objective criteria for such assignments since, according to the Bayesians, all probabilities are subjective or personal. Thirdly, a condition for the applicability of probability is chance or randomness—a condition absent from every fairly well organized body of knowledge. Finally, it is doubtful whether many people do learn from experience in any rational way, let alone in accordance with Bayes' theorem—particularly when probabilities are involved. In short, subjective probability does not solve the problem of measuring the state of knowledge nor, a fortiori, its evolution.

(Compare the claims of subjectivists with the actual way probabilities are used in science and technology. Here there are two possibilities: either one knows the probabilities of interest or one does not. If the former, it is either because of measurement or of computation, and in either case the posited probabilities are *objective* in the senses that they quantitate objective possibilities and that they are subject-free, i.e. public. If on the other hand the scientist or technologist is faced with total ignorance concerning the probabilities of alternative possibilities, or the distribution of a continuous stochastic variable, he adopts either of the following strategies. Either he abstains from assigning any values to the probabilities or the densities in question, or he makes hypothetical assignments, watches for the logical consequences—and makes the necessary corrections. He realizes that any such hypothetical assignments—e.g. equiprobability—are *hypothetical* and part of the particular model not of the mathematical theory of probability. On the other hand the Bayesian statistician makes arbitrary assignments of prior probabilities without regard to any matters of fact: he defies the scientific method. To the scientist or technologist any probabilities are probabilities of states or events, and therefore they (*a*) depend on the nature of the thing concerned and (*b*) must be either measured or conjectured regardless of the verdict of the subjectivist probabilist. More in Bunge (1981c).)

Finally, science and technology watchers and administrators have developed objective though ambiguous indicators of the volume and rate of growth of public knowledge. In fact there is an entire industry devoted to such measures as well as specialized periodicals, such as *Citation Index*, and annals such as *Science Indicators.* No doubt, such "scientometric" indicators do measure the *volume* of publications and their *diffusion.* Whether they also measure the volume and growth of *knowledge* is doubtful. Merely counting publications does not discriminate between original research and rehash, between deep insight and shallow information, or between revolutionary proposal and routine stuff. And counting the number of times a publication is cited indicates popularity rather than intrinsic value. The difference is important for, as a rule, the newer and deeper a contribution, the harder it is to understand and therefore the least popular it is bound to be at least for a while.

2. MODES OF KNOWLEDGE

2.1. *Basic Categories*

Classical epistemology dealt with the lone knowing subject and his knowledge, and did so in a static manner: it minimized the cognitive process and ignored its interaction with social behavior. The epistemology we are sketching deals instead with animals engaged in learning processes in their environment, and regards knowledge as a phase in those processes. Thus it interprets the conventional expression 'X is known' as "There is at least one animal of species S who has learned X under circumstances Y". At the same time we recognize the need for studying the "product" of the cognitive process regardless of the idiosyncrasies of the learning subject and her environment—i.e. the need for the study of knowledge.

Now, 'knowledge' is an omnibus word: it covers all learned skills (not the inherited ones though) and all manner of ideas. Indeed knowledge can be:

(i) *sensory-motor*—e.g. knowing how to walk (or limp) or how to type (however clumsily);

(ii) *perceptual*—e.g. tasting a lemon as bitter (or otherwise), or seeing a knoll of grass as green (or purple);

(iii) *conceptual* or *propositional*—e.g. knowing that the Earth revolves around the Sun (or conversely), or that the heart pumps the blood (or is the seat of the soul).

To put it into classical terms: knowledge can come from any of three "sources"—action, perception, or ideation. Therefore radical pragmatism ("Action is the only source of knowledge") does contain a grain of truth. So do empiricism ("Perception is the only source of knowledge") and rationalism ("Cogitation is the only source of knowledge"). By the same token each of these epistemologies is mistaken in excluding the other two "sources" of knowledge. (Caution: memory, traditionally regarded as a source of knowledge—about the past—can at most distort acquired knowledge, never generate it.)

The above distinction among the three knowledge categories should not blind us to the fact that they are interrelated. Thus conceptual knowledge can improve motor skills and perception: the trained hand (or eye or ear) performs better than the naive one. Besides, all three are simultaneously present in many cognitive activities, such as drawing and writing.

Note also that knowledge of any type may be correct or incorrect to some degree. Thus most of us would be unable to make anatomical drawings like those of Ramón y Cajal, or even Vesalius or Leonardo; most city dwellers fail to perceive (sense and identify) the sounds of nature as acutely and precisely as country folks; and all of us know as many falsities as truths. (And this trivially, in the sense that whoever knows the proposition p knows *eo ipso* not-p, as well as in the deeper sense that much of our conceptual knowledge is at best approximately true.) In short, knowledge is not necessarily correct. Nor does knowledge involve belief: I know many "things" (e.g. ideological doctrines) in which I do not believe, and I grant beforehand that some of my beliefs are false. Therefore the standard definition of "knowledge" as true belief is inadequate. More in Section 3.2.

A second useful distinction is that between *self-knowledge* (or knowledge of oneself) and *other-knowledge* (or knowledge of something other than oneself). Both can be obtained directly through perception, and self-knowledge also from feeling and emotion. (Feeling and emotion are not cognitive operations, but our awareness of them is a cognitive item.) They can also be known indirectly, namely by cogitation. And either mode supplements, checks, and corrects the other.

The ancient and wise injunction *Know thyself* can now be rendered more precise, to read *Know thy brain and thy social behavior.* To be sure one will never get to know much about oneself without the help of neuroscientists, psychologists, and social scientists. However, one can learn a lot about oneself by watching one's feelings and desires, modes of thought and behavior patterns—all of which *is* part of knowing one's brain and one's social behavior. A scientific study of all that will yield more detail and, in particular, a deeper understanding of why we feel, desire, think and behave the way we do.

Other-knowledge can be of things (natural, social, or artificial) or of constructs (concepts, propositions, theories, or proposals). And the latter may, though it need not, involve the former. Thus knowing a reasonably true atomic theory implies knowing something about atoms. On the other hand knowing some differential equations does not imply knowing any real things except the inscriptions representing such equations. (Rather on the contrary, we need some knowledge of differential equations in order to get to know something about atoms in any depth.)

We emphasize that knowing facts is not the same as knowing propositions about facts—a distinction that epistemic logic fails to draw. When perceiving fact f, or performing action f, we may gain *direct* knowledge of f; this knowledge is not necessarily propositional. Hence we are justified in asserting "I know f", rather than "I know that the proposition p describing fact f is true". On the other hand when learning about facts in books, or when working out scientific theories, we have *indirect* or propositional knowledge of them. These two modes of knowing, κ_1 and κ_2, compose as shown in the following self-explanatory diagram:

$$\begin{tikzcd}
\textit{Facts} \arrow[r, "\kappa_1"] \arrow[dr, "\kappa_1 \circ \kappa_2"'] & \textit{Propositions} \arrow[d, "\kappa_2"] \\
& \textit{Propositions}
\end{tikzcd}$$

Table 2.1 exhibits some types of knowledge distinguished by subject matter or referent.

A third useful distinction is that between first and second hand knowledge. *First hand* knowledge is acquired by personal experience, in particular research. *Second hand* knowledge is knowledge about first hand knowledge: it is communicated by word of mouth, the media, textbooks, or review articles. This partition of knowledge items does not involve any

TABLE 2.1
Types of knowledge with regard to subject matter or referent

Type	*Referent(s)*	*Field(s)*
Formal	Constructs	Logic, mathematics, semantics
Factual	Things, facts	Ordinary knowledge, arts, crafts, factual science, technology, ontology
Empirical	Experience	Ordinary knowledge, arts, crafts
Moral	Right conduct	Ethics, sociology, art
Epistemological	Knowledge	Epistemology

valuation. First hand knowledge may be worthless—e.g. I know there is a tiny inkspot on this sheet of paper—and second hand knowledge may be high grade—e.g. an insightful critical review of recent work on the effectiveness of psychotherapy. Knowledge of both types is necessary, for nobody can learn everything first hand. But we should not allow second hand knowledge to get the upper hand—as it did during the Middle Ages.

A related distinction is that between private and public knowledge. We may say that animal b has *private* (or privileged) knowledge of X if and only if there is nobody except b who has knowledge of X. Otherwise, i.e. if an item of knowledge is shared among at least some members of a society, it is *public* in that society. These are just definitions. Finding out whether a given bit of knowledge is public in some society calls for empirical investigation by psychologists, anthropologists, social psychologists, sociologists of knowledge, or historians of culture. Mere linguistic analysis will not suffice, notwithstanding those philosophers who claim that 'p is known' is an obvious example of public (or impersonal) knowledge (e.g. Williams, 1972). The fact that someone holds something to be known, without caring to mention who knows it, does not entail that such knowing can be impersonal or suprapersonal: it only indicates that the individual who makes that claim does not care to tell us who knows that particular item. The ordinary language expression 'p is known' must be analyzed into 'There is some animal that knows p'. Likewise when we say that 'x attracts (or repels)' we presuppose that there is something, other than x, that is attracted (or repelled) by x. *Moral*: Language is not a tool of analysis but an object of it.

There are two kinds of private and two kinds of public knowledge. Private knowledge can be knowledge of one's own states, in particular brain states, or secret knowledge, i.e. knowledge kept "classified". The former is direct in the sense that the subject has access to it without the help of anyone else. On the other hand getting hold of a secret concerning somebody or something requires data supplied by others—deliberately or unwittingly. However, knowledge of the two kinds can, in principle, be made accessible to others, i.e. rendered public in part. Surely your headache will not be felt by anyone else, but it may be guessed from your grimaces or from objective indicators such as, perhaps, the rate of endorphin synthesis in your brain. And, if physiological psychology continues to advance at its present pace, conceptual privacy may soon vanish. So much for the claim that all mental events are private.

Public knowledge too comes in two varieties: *ordinary* and *specialized.*

The former is shared by (nearly) all (adults) of a given community (at a given time), whereas specialized knowledge circulates (at a given time) only within very special learning communities, such as those of masons or mathematicians. The philosophers who possess no specialized knowledge tend to revere ordinary knowledge, claiming that it is sufficient for all purposes, or at least that it should be regarded as the Supreme Court of Cognitive Appeals, for being solidly anchored in experience and untainted by prejudice and theory. Both claims are wrong. The mere existence of mathematics, science, technology, and philosophy shows that common sense was found insufficient, for they start off where the latter stops or founders. Nor is ordinary knowledge as solid as it looks: it is mostly superficial, largely inaccurate, and remarkably incomplete. And in any event ordinary knowledge is shaped and updated not only by personal experience but also by the entire culture, as it feeds off the stale crumbs of science, technology and philosophy that may come within its reach. (By the way, there is no such thing as "common sense philosophy". This is a contradiction in terms: whatever is commonsensical is uneducated and therefore likely to be uncritical, shallow, and obsolete—in short, unphilosophical.)

The second variety of public knowledge, namely specialized knowledge, is supposed to be a result of specialized practice or of inquiry by individuals or teams possessing a special background and special skills. It comes in a variety of grades, from expertise (e.g. in cooking or managing or computing) to scientific, technological or humanistic knowledge. Here again, truth and usefulness are irrelevant: whereas some artisanal know-how is true and effective, some scientific (or technological or humanistic) knowledge is false and useless.

Plato's dichotomy between *doxa* (groundless opinion) and *episteme* (demonstrable science), though widely accepted, is inadequate. An electronic technician, though no scientist or technologist, uses well tested scientific and technological hypotheses, data and procedures, whereas a scientist cannot help hazarding speculations, and a technologist cannot avoid making recommendations on the basis of incomplete knowledge topped with intuition and judgement. The relevant distinction is not that between *episteme* and *doxa* but that between knowledge that is, or is not, a result of scientific (or technological or humanistic) research, as well as a possible subject of further investigation. In other words, the important distinction is that between research fields and belief systems. We shall return to this point in Section 4.1.

Another untenable dichotomy is that between *pure* and *practical* knowledge. Plato and Aristotle emphasized this distinction, praised the former and deprecated the latter. The pragmatists deny it, asserting that all "knowing is for the sake of doing" (Lewis, 1946, p. 3). However, some knowledge starts by being pure and remains so: it was acquired for its own sake and never finds application. Thus we do not "do" anything with cosmology, and nothing good has so far come out of the study of history—yet we cherish both and are proud of a culture that renders the cultivation of such disciplines possible and even desirable. However, it is hardly possible to forecast that a given piece of knowledge will remain pure forever. Electromagnetic fields, atoms, genes, fossils, brains, and many other things were initially studied with no practical purposes in mind, whereas nowadays entire technologies are concerned with using or manipulating them. In short, the pure-practical distinction does not amount to a dichotomy.

2.2. *Tacit and Explicit Knowledge*

Only some of our knowledge, whether sensory-motor, perceptual or conceptual, is conscious or explicit and expressible in some language: much of it is unconscious or tacit in the sense that we possess it without our knowing or without our being able to express it in any language. Thus craftsmen and artists, technologists and scientists, know how to do a number of "things" that they could not put into words. (The great Euler is reputed to have said that his pencil knew more than himself.) And any honest technology transfer deal stipulates that the customer will get not only artifacts and the corresponding operating and service manuals but also experts capable of teaching by example how to handle and service the system—precisely because not everything can be put into manuals.

A possible definition of the above concepts is this: If subject s knows p, then (a) s has *explicit* knowledge of p if and only if s also knows that s knows p or knows how to express p in some language; (b) otherwise s has *tacit* knowledge of p.

The difference has been explicitly drawn by a number of philosophers, in particular James (1890 I), and Russell (1918), who distinguished between *knowledge by acquaintance* (or *tacit* or *procedural* knowledge) and *knowledge by description* (or *explicit* or *propositional* knowledge). Polanyi (1958) emphasized the importance of tacit knowledge (which he called 'personal'); so did Oakeshott (1962). And both held that tacit knowledge is superior to explicit knowledge. This is a curious dogmatic statement to make for, by definition, tacit knowledge is ineffable and therefore hardly subject to examination, in particular evaluation.

Both tacit and explicit knowledge can stay private or become public, depending on the knowing subject, the degree of advancement of the field of knowledge, and society. However, it is obvious that explicit knowledge, for being expressible rather than ineffable, is far easier to share—provided there are no social barriers to information. (Remember that nearly every society regards knowledge of certain kinds as esoteric, or accessible only to a small group of initiates, be they priests, politicians, or intelligence agents.) But all this is rather obvious and therefore philosophically unproblematic.

A problem of interest to the cognitive sciences and epistemology is whether there are neurophysiological conditions that impose the dissociation between knowing how and knowing that or whether, with some effort, we could render explicit every bit of tacit knowledge. The scientific investigation of this question has only started. We have learned that amnesic patients can learn certain skills even if they cannot remember the information gained by using them. Thus patients unable to remember day-to-day events (i.e. who have lost short term memory) can be taught mirror reading without their remembering which words they had read (Cohen and Squire, 1980). This finding suggests that the dissociation between tacit and explicit knowledge is quite deep-seated.

We still do not know, then, whether there must always remain a residue of ineffable knowledge. But at least we seem to have learned that the right way to approach this problem is through scientific investigation rather than by resorting to a ready-made doctrine (such as intuitionism, rationalism, or empiricism). The problem is not one of mere academic interest: it is also of interest to educationists and to experts in the transfer of technology. Indeed, both groups would like to know whether every bit of knowledge can eventually be poured into a manual or a tape.

The tacit (or procedural)/explicit (or propositional) distinction could be used to discuss the problem whether computers can be said to know anything. It might be argued that they can acquire know-how but not know-that: that their knowledge is purely procedural—e.g. that they know how to add integers but not what addition or integers are. However, if we stick by our psychobiological approach we cannot admit that computers know anything because, having no nervous systems, they cannot experience plastic changes in them. All we can say is that computers are in some respects similar or analogous to knowing subjects, just as jets are in some respects similar to birds. These analogies are all-important for technology but of very little philosophical interest.

The tacit-explicit distinction intervenes also in current discussions in psycholinguistics. Empiricists claim that knowing a language is an instance of know-how in the same category as bicycling or hammering. Extreme rationalists, on the other hand, hold that knowing a language, like knowing anything else, is a case of know that (or propositional knowledge) in the same category as knowing that jets are fast or that the day has 24 hours. And Chomsky (1980) holds moreover that the basic grammar of a language is an inherited endowment somehow modified by experience and reason, perhaps in much the same way as songbirds improve by practice and imitation the songs they inherited genetically.

We will not repeat here what we held in Ch. 1, Section 3.1 about the inheritability of neurophysiological features that make it possible for a human to learn a language under suitable social circumstances. But we must pose the following problems. First, is it not likely that the young child's knowledge of language is a case of know-how in the same category as his knowing how to play with toys, and that such know-how turns into know-that only with formal education? Second, if knowing a language is to master the rules for generating grammatical sentences, and if such rules are innate or at any rate they are learned long before the speaker can form a notion of a rule, then are we really talking about *rules* proper, i.e. prescriptions for doing something, or rather about dispositions or even laws?

Whether the grammatical rules are inherited or learned as bits of know-how, they may be objective *patterns* ingrained in the brain rather than the constructs we learn at school. Thus knowing how to speak English amounts to having acquired the pattern of generating English sentences without necessarily *knowing* it, much as we learn to walk or play peekaboo. Likewise planets may be said to possess the laws of planetary movement without knowing them. Hence the expression 'knowledge of a language' is misleading if 'knowledge' is understood as being always conscious: it is mostly a matter of know-how rather than of know-that. Neither children nor illiterate adults nor the apes who have learned to manipulate human linguistic signals have explicit knowledge of language, hence they cannot make use of any grammatical rules in order to avoid or correct grammatical mistakes. They have only tacit knowledge of a language. Formal learning can bring this knowledge to the fore, i.e. transform tacit into explicit knowledge, patterns into rules proper.

Now two cautions against certain rather common confusions. The first is that tacit knowledge should not be equated with private or privileged knowledge (Section 2.1), for the former can often be shared or transmitted, if not verbally by example. And explicit knowledge may but need not be public: it may be kept secret. All public knowledge is or may be rendered explicit but the converse is false.

Our second caution is against mistaking public knowledge for objective knowledge, as explicitness and publicity are only necessary conditions for objectivity. Indeed we define objectivity as follows. First the general concept, applicable to both factual and formal knowledge. Let p be a piece of explicit knowledge. Then p is *objective* if and only if (*a*) p is public (intersubjective) in some society, and (*b*) p is testable (checkable) either conceptually or empirically. And now the strong or special concept of objectivity. Let p be a piece of objective knowledge. Then p is *factually objective* if and only if p has a factual reference (i.e. is about actual or possible facts).

Note again the difference between objectivity and intersubjectivity, a distinction that most epistemologists do not draw. (No subjectivist epistemology has any use for it.) Magical rules and religious dogmas may be intersubjective in a given society but they are not objective in the above methodological sense. As for the difference between objectivity *lato sensu* and *stricto sensu*, note that mathematical theorems are objective in the former sense but not in the second, whereas the statements made in science and technology are supposed to be objective in the narrow sense. (This point is of course being controverted in the social sciences, in particular in history. We shall take it up again in Vol. 6.) Finally, note that truth is not involved in objectivity. A statement may be objective and false, or true and non-objective. For example "$2 - 1 = 0$" is objective and false, and "I am not sure I'll ever finish reading this book" may be true but not objective. More on objectivity in Ch. 15, Section 2.2.

3. BELIEF

3.1. *Belief System*

In line with our biological approach to mind we construe beliefs, doubts and disbeliefs as brain processes of the mental kind. Like any other brain process of this kind, a belief has a genesis and it can change, even disappear, under the impact of other bodily (in particular mental) processes.

A possible characterization of belief and its cognates is as follows (modified from Bunge, 1980a, Postulate 7.3). Let b be an animal, u and v thoughts, and τ a time interval. Then:

(i) *b entertains* thought u during τ iff b has psychons that think up u, as well as other thoughts related to u, during τ;

(ii) *b believes* thought u during τ iff b entertains only u during τ (i.e. the psychons for other thoughts, related to u, are not activated during τ);

(iii) b is *doubtful* (or *undecided*) between thoughts u and v during τ iff b entertains both u and v alternately during τ, i.e. in such a way that the activation of the psychon for u inhibits that for v and conversely, i.e. cyclically.

Being brain processes, believing, doubting and disbelieving can in principle be investigated empirically (by neuropsychology). And, because the holding of certain beliefs controls social behavior, belief can also be investigated by social scientists. In fact it is being investigated by social psychologists and other workers. Such empirical approaches are of course at variance with the a priori approach adopted by the philosophers interested in the "logic" of belief (e.g. Hintikka, 1962).

Like any other brain processes, beliefs come in different strengths. We may grade them between -1 (maximal disbelief or rejection) and $+1$ (maximal belief or acceptance); and we may assign the value 0 to any thought that we are as strongly inclined to accept as to reject. More precisely, call B_b the collection of thoughts that animal b can initiate at time t, and T the set of all instants. We conjecture that there is some (presently unknown) function w_b from the cartesian product $B_b \times T$ into the unit real interval $[-1, 1]$, whose value $w_b(x, t)$, for x in B_b and t in T, is the *weight* or *strength* of x for b at time t.

Presumably, the values of w_b for different thoughts and times could be determined empirically. Such empirical research would show that belief strengths change in the course of time, occasionally in a dramatic fashion. If we can determine the collection B_b of thoughts of animal b, together with their respective weights, we have *eo ipso* determined the *belief system* of b, which we define as $\mathscr{B}_b = \langle B_b, w_b \rangle$.

To be sure, we have not determined the weight functions w_b but have merely expressed the hope that they may one day be determined empirically. As a matter of fact beliefs are already being subjected to experimental investigation, which is not surprising considering that they are brain processes rather than inmates of an immaterial mind. Being brain processes they can be influenced by suitable cognitive stimuli as well as by the subject's own ongoing brain activity. For example, a subject led to believe that she had been touched with the leaves of a harmful plant (e.g. poison ivy), while in fact she had been touched with a harmless plant, may develop flushing, itching, and erythema (Barber, 1978). The findings of social psychologists, on the way belief can be manipulated, are, if anything, even more dramatic: see the classical studies of Sherif (1936) and Asch (1952).

Another problem concerning belief that has been investigated experimentally is the discrepancy between what we believe and what we *say* we believe. For example, in a given English speaking community most of the women may pronounce 'student' while claiming that they always say 'stjudent', which is the socially accepted pronunciation; men behave in the dual manner. (It is not that all those subjects lie, except to themselves: they may not be aware of the discrepancy.) In any case of this kind it is the experimenter's task to fill out the following matrix:

	Professing to believe or do *p*	Not professing to believe or do *p*
Actually believing or doing *p*	*a*	*b*
Actually not believing or doing *p*	*c*	*d*

with $a+b=1$ and $c+d=1$. If all subjects are perfectly truthful (to the experimenter and to themselves), $b=c=0$. In actual fact this situation is seldom encountered. I.e. most of us do not quite know what we believe or do—which is one more argument against the definability of knowledge in terms of belief.

To return to the belief weights w_b. Since beliefs are brain processes, in principle they might be assigned objective probabilities conditional on experience and circumstances. One might speak of individual b holding belief u at time t, with probability p (or of the individual's propensity or disposition to believe u at t). In such case we would set $w_b(u, t) = 2P_{bt}(u) - 1$.

And, of course, one might then introduce the conditional probability of a given individual switching from one belief to another. But, as noted, such probabilities must be objective, for it makes little if any sense to ask an experimental subject to reply to questions of the form 'What subjective (or personal) probability do you attribute to that proposition (e.g. the Schrödinger equation): 0.7, 0.8, 0.9, or 1.0?'. And they must be conditional: a person accustomed to evaluating beliefs critically is less gullible than one who has received a dogmatic education, and so the probability of his acquiring a belief in the ghostly is smaller than the average.

The preceding considerations do not carry over, without further ado, to groups of animals, such as professional groups or social classes. To be sure ideologists and even some social scientists sometimes claim that the bourgeoisie, or the proletariat, or some other social group, holds this or that belief, or has this or that belief system. From a psychobiological viewpoint this claim is false, because social groups do not have collective brains. (On the other hand the claim does make sense in the context of objective idealism, where there is a body of ideas floating around, which some individuals can grasp albeit only imperfectly.) However, the notion of group belief, or belief system of a group, can be rendered rigorous and useful in the following way. Form the intersection of all the beliefs of the individual members of a group, and the family of their belief strengths: i.e. collect all the beliefs that the group members share (hold in common), and assign each its weight. The result is this: The *belief system of group G* is the structure $\mathscr{B}_G = \langle \bigcap_{b \in G} B_b, \{w_b \mid b \in G\} \rangle$.

The intersection of the beliefs of the members of a group of animals of the same species is never empty. But if it is small, as it happens with every wide group and every intensely divided society, the concept of group belief is not of much use, and the social scientist is forced to subdivide the given group into more homogeneous groups. But when the intersection is large, as is the case with homogeneous groups, we can speak of *prevailing opinion*, or *climate of ideas*, or "spirit of the times" (*Zeitgeist*)—provided we do not forget that such prevailing opinion is neither held by a single individual nor by a ghostly collective mind, but is a construction out of individual belief systems.

We should not underrate the weight of prevailing opinion on inquiry. Indeed such prevailing opinion may be favorable, unfavorable, or indifferent to a given line of inquiry (or research project). If favorable, "it" will encourage the project to the point of overlooking blemishes and discouraging competing projects. If unfavorable, "it" will exaggerate shortcomings and erect obstacles—which may either kill the project or force the investigators to be more ingenious, resourceful, careful, and critical. In either case, whether favorable or unfavorable, prevailing opinion is double-edged. (It works like a two loop control system, one of the elements providing negative, the other positive feedback.)

We are all sensitive, in some degree or other, to prevailing opinion and group pressure. Hence everyone is more or less suggestible and gullible, in addition to taking the good faith of others for granted. This makes us all more or less easy preys to opinion manipulation by teachers and preachers, politicians and ad-men. Contrary to widespread opinion, not even scientists are immune: the chemist who would not accept uncritically reports on success in synthesizing a certain molecule may fall in for homeopathy, psychoanalysis, or monetarism. A nationwide Gallup poll conducted in 1978 to assess the degree to which Americans believe in the paranormal and the ghostly gave the following results: 57% believe in UFOs, 54% in angels, 51% in telepathy, 39% in devils, 37% in precognition, 29% in astrology, 24% in clairvoyance, and 11% in ghosts. At about the same time it was found that belief in the paranormal is much higher among university students than among the public at large—a sad comment on the failure of the university to educate (Greenwell, 1980).

Why do so many people hold such groundless beliefs at a time when we are supposed to embrace science and technology? Presumably because some of those ideas seem to explain easily a great many phenomena, others are comforting, still others promise instant knowledge or power, and most of them have been received uncritically at an early age from the previous generation along with fairy tales and religious dogmas. In all such cases belief is supported not by argument or theory, let alone empirical evidence, but by tradition and a natural disposition to receive whatever one is given, as well as by the wish to secure cheap protection. (It is said that all risky and highly competitive professions are ridden with superstition. Sailors and airline pilots, truck drivers and stock brokers, soldiers and prostitutes often wear amulets.) The critical attitude and scientific inquiry are still recent and uncommon. Still, interestingly enough, few believers in those superstitions act consistently on the strength of such beliefs. For example, believers in astrology seldom choose their profession by consulting their horoscopes. There is always a gap between belief and action, principle and deed.

Experience is no deterrent to superstition because it can be interpreted in terms of the latter. On the contrary, we often acquire superstitious beliefs from having experienced certain coincidences and from our unwillingness to admit chance (a very modern idea that has not yet seeped down). This has been suggested by a classical experiment on the origin of superstition in pigeons. Instead of reinforcing "good" behavior, Skinner arranged for food to be delivered to them at regular intervals regardless of the animal's behavior. (I.e. he gave them adventitious reinforcement, which is what we do with our children when feeding them cookies and TV when they return from school regardless of their grades.) It turned out that one bird was conditioned to turn about the cage, another thrust its head into the upper corners of the cage, a third developed a tossing response, and two birds a pendulum motion of the head and the body. "The conditioning process is usually obvious. The bird happens to be executing some response as the hopper [food dispenser] appears; as a result it tends to repeat this response" (Skinner, 1948, p. 168). True, there is some evidence that such superstitious behavior does not occur in rats unless they have hippocampal lesions (Davenport, 1979). Still, chance coincidence must play some role in the formation of groundless beliefs even among higher vertebrates. And among humans docile reception of traditional belief is often socially rewarded.

Uncertainty is unpleasant to most of us, and this is why dogma is so popular. Indeed the adoption of dogma removes uncertainty, for a dogma, or at least a clever one, has ready-made answers to all possible problems of a kind. On the other hand inquiry, particularly in science, technology, and the humanities, creates new uncertainties for every uncertainty it removes, as every research problem ends up, in the case of success, by posing new problems while solving some. Thus, inquiry does not decrease uncertainty. But of course this is not the unsettling and paralyzing uncertainty produced by the failure of dogma: it is the stimulating uncertainty that promotes further inquiry. (If only for this reason the information-theoretic model of the growth of knowledge is inadequate, for it equates such growth with information gain, which would in turn be the same as decrease in uncertainty.)

Having condemned dogmatism let us now sound a warning against radical or systematic skepticism. If we were skeptics all the time we would abstain from making choices and decisions, as well as from acting on them. Consequently we would soon die—ignorant to boot. Radical skepticism, in particular skepticism about all our own perceptions and thoughts, has no survival value. Here as elsewhere a middle course is advisable, namely *methodical skepticism*.

3.2. *Belief and Knowledge*

We saw in the preceding section that we do not always know *that* we believe or *what* we believe. And sometimes we profess to believe "things" we really do not know anything about, such as life after death, perfect rationality, and spotless democracy. Some people go even further: thus mystics and psychoanalysts claim that their doctrines must be believed in order to be understood. (Remember Isaiah's *Credo ut intelligam*—I believe in order to understand—and the psychoanalytic requirement that every practitioner be subject to analysis.) A few have gone to the extreme *Credo quia absurdum* (I believe because it is absurd). In all these cases belief is recommended as prior to understanding and testing. Believers in the supernatural or the paranormal are faithful to this maxim: rather than believing science they prefer to believe what flies in the face of evidence and reason. Thus Kierkegaard, realizing that Christian dogma is absurd, stuck to it and declared existence itself to be absurd.

The uneducated believes what he knows and even what he does not: he is basically a dogmatist, only occasionally an inquirer. The educated knows what he believes and does not quite believe what he knows: he is basically a skeptic, though seldom a systematic or radical one. In short, belief is not a condition for knowledge. On the contrary, knowledge is a condition for rational or justified belief (or disbelief or suspension of belief). *Knowledge, then, is not a kind of belief.* Yet, Russell (1948) and many other epistemologists have defined knowledge as a special kind of belief, namely as *justified true belief*. (More explicitly: *s* knows that *p* if and only if *s* believes *p*, *p* is true, and *s* is completely justified in believing *p*. See Lehrer (1974) for a fuller statement of the belief analysis of knowledge.) This view is refuted by all the cases of people who profess to believe what they really do not know, and disbelieve a number of "things" they know. A more sophisticated counter-example is this: The disjunction of an arbitrary proposition with a tautology is true, so according to the belief theorists we should be justified in believing it. However, such a disjunction cannot pass for knowledge, particularly if the nontautological component is arbitrary or indeterminate. Besides, according to the definition of knowledge in terms of belief, most scientific statements would not qualify as bits of knowledge because they are at best partially true (see Ch. 12, Section 1.1). Nor would abstract theories, such as set theory and group theory, qualify, for they are neither true nor false. Being inapplicable to the most interesting bits of knowledge, the belief analysis of knowledge is itself unbelievable.

A rational person will make some effort to know a proposition before believing or disbelieving it: his beliefs are special cases of his knowledge not the other way round. Consequently instead of defining knowledge as justified belief, a rational person will define justified belief in terms of knowledge. He may accept the following definitions:

(i) *s believes* $p =_{df}$ s knows p & s gives assent to p;
(ii) *s is justified in believing* $p =_{df}$ s knows p, and p is approximately true or has a model;
(iii) *s is justified in doubting* $p =_{df}$ s knows p and s knows of no grounds to assign p a definite truth value or a model;
(iv) *s is justified in disbelieving* $p =_{df}$ s knows p and s knows that p has been refuted or has no models.

The problem of the relation between knowledge and belief can be approached either descriptively or normatively. The descriptive study of cognitive attitudes, such as belief and disbelief, certainty and doubt, is a legitimate field of inquiry if conducted scientifically, as part of psychology. It is illegitimate otherwise, i.e. if conducted without regard for empirical testing. A normative investigation of cognitive attitudes, i.e. one serving as a basis for a rational theory of decision and action, should be empirical as well: it should be possible to check whether the norms it proposes are efficient, i.e. conducive to successful action. Clearly, any such normative theory must presuppose some notion of knowledge and some notion of truth, for othewise it may consecrate groundless belief, i.e. dogmatism.

One may go farther and demand that such a normative theory of belief contain the axiom that credences, or degrees of belief, be numerically equal to the corresponding degrees of truth. (That factual propositions are true or false to some degree or other is known to all scientists and technologists. For the notion of partial truth see Ch. 12, Section 1.1, and Appendix 3.) More precisely, we postulate that, if p and q are distinct propositions, and the degree of truth of p given q (as a baseline) is v, then the rational credence of p given (assuming) q should equal v. (On the other hand if p and q are the same, i.e. p is its own ground, and thus groundless except possibly for its consequences, then p should be assigned no credence at all.) A consequence of this postulate is that the truest propositions are assigned the greatest rational credences. Another is that, if p is true, then not-p is incredible. A third is that all the consequences of a true proposition are maximally credible. Clearly, any theory containing such axiom is normative not descriptive: it is a theory of rational or grounded belief.

The standard theories of knowledge and belief, namely epistemic logic and doxastic logic, first proposed by Hintikka (1962) and now active subjects of research (cf. Lenzen, 1980), are of a totally different kind. In fact they are a priori (i.e. "logics") and they do not use the concept of partial factual truth, so they are hardly relevant to scientific, technological or humanistic knowledge. To begin with, the basic concepts of epistemic logic, namely those of belief, certainty, and knowledge, are taken from ordinary knowledge. Surely they are symbolized, but symbolization is no guarantee of truth or depth. For example, the principle "If s believes p, then possibly p" remains hazy as long as neither belief nor possibility are adequately elucidated. Hintikka (1969) does elucidate the notion of possibility in model-theoretic terms: the outcome is the so-called possible worlds analysis of knowledge. However, the possibilities concerned are conceptual not factual, whereas in the case of factual knowledge we are interested in real (nomological) possibility. (See Vol. 3, Ch. 4, Section 2.) Secondly, epistemic logic is not interested in the grounds for holding beliefs—such as empirical evidence and harmony with previously confirmed hypotheses—and so it makes no contact with factual truth. Consequently it tolerates arbitrariness and irrationality, and so it is not a tool for examining beliefs with a view to justifying or criticizing them. Thirdly, epistemic logic deals with isolated propositions and arbitrary sets of such, which is not the usual case in science, technology, or the humanities. In sum, epistemic logic is not an efficient tool for analyzing knowledge, and thus belongs neither in descriptive nor in normative epistemology. It is just a *jeu d'esprit.*

It does not follow that Popper is justified in holding that belief and the study of it are unimportant. On the contrary, we live and die by our instincts and beliefs. In particular all our inquiries, valuations and conscious actions are guided or misguided by beliefs. Thus, we believe in the value of scientific research to know the world. Only, knowledge is not a special kind of belief, and therefore epistemology is not included in the study of belief; moreover, the latter belongs in psychology.

4. INQUIRY

4.1. *Field of Knowledge*

Knowledge comes in degrees, and so does its dual, i.e. ignorance. Unlike knowledge, total ignorance cannot be assigned a brain activity of a definite type: it is the absence of such activity. On the other hand realizing, like Socrates, that one ignores something, is a bit of knowledge and therefore a brain activity. There are two ways of overcoming ignorance. One, cheap but illusory, is to embrace uncritically some belief; the other, onerous but effective, is to engage in inquiry in order to obtain some first hand or second hand knowledge.

Consider the main steps in a typical process of inquiry.

Step	*Events in inquirer's brain*
1	Considers problem p.
2	Inquires about theoretical or empirical setting of p.
3	Thinks up possible solution s to p.
4	Makes preliminary check of s
5	Concludes that s is inadequate.
6	Searches for alternative solutions to p.
7	Comes up with candidate t.
8	Makes preliminary check of t.
9	Concludes that t may be adequate.
10	Searches for counterexamples to t.
11	Having found none (or having explained them away) declares t to be a plausible solution to p.
12	Decides to investigate t methodically.
13	Formulates tactical plan for investigating t.
14	Writes up proposal for research grant.
15	In the process finds new data or reasons relevant to problem p.
16	New information suggests reformulating or even radically changing p.
17	Above change forces modification of original research plan.
18	Writes up final proposal.
19	Works on problem while proposal is being considered.
20	Takes cognizance of rejection of research project on the ground that problem is worthless or not original, or that proposed method or suggested solution is inadequate or too original.

Inquiry is a particular kind of cognitive process: it is *directed cognition*. In fact inquiry stars with some problem found in a given fund of knowledge, it employs definite means, and aims at finding some thing or idea that may solve the problem. Inquiry takes then at least five items: a group of inquirers, an incomplete fund of knowledge, a set of problems (gaps in that fund), a set of research tools (conceptual or material), and a set of goals (cognitive or practical). Successful inquiry builds up an epistemic field as it proceeds: logic or geometry, physics or biology, engineering or computer science, and so on. Since we shall make much of the concept of an epistemic field in this volume and the next, we may as well proceed to give a definition of the concept. (See Törnebohm (1979) for a different though congenial concept.)

The ten-tuple $\mathscr{E} = \langle C, S, D, G, F, B, P, K, A, M \rangle$ is an *epistemic field* (or *field of knowledge*) at a given time iff, at that time,

(i) C *is a system* composed of persons who have received a special training, hold strong information relations amongst them, and initiate or continue a tradition of belief or inquiry;

(ii) S *is a society* capable of supporting and encouraging, or at least tolerating, C;

(iii) the *domain* or *universe of discourse* D of $\mathscr{E}$ is the collection of objects of the inquiry, i.e. the reference class of the specific concepts of $\mathscr{E}$;

(iv) the *general outlook* or *philosophical background* G of $\mathscr{E}$ is composed of (*a*) an *ontology* or view on the nature of things, (*b*) an *epistemology* or view on the nature of knowledge, and (*c*) a *morality* concerning the proper ways of acquiring, diffusing and utilizing knowledge;

(v) the *formal background* F of $\mathscr{E}$ is a collection of logical or mathematical theories taken for granted in the course of the inquiry;

(vi) the *specific background* B of $\mathscr{E}$ is a collection of propositions and procedures, other than F, drawn from other epistemic fields;

(vii) the *problematics* P of $\mathscr{E}$ consists of problems concerning the nature, value or use of the members of D as well as problems regarding other components of $\mathscr{E}$;

(viii) the *fund of knowledge* K of $\mathscr{E}$ is a collection of propositions and procedures obtained by members of C at previous times;

(ix) the *aims* or *goals* A of $\mathscr{E}$ are of course the target (cognitive, practical or moral) of the specific activity of the C's;

(x) the *methodics* M of $\mathscr{E}$ consists of all the general and special methods utilizable in $\mathscr{E}$.

The first three components of this ten-tuple will be said to compose the *material framework* of the epistemic field in question. (Clearly this is a misnomer in the case of formal science, the humanities and theology, all of which deal primarily with conceptual objects not material ones. However, the expression has the virtue of reminding us that at least the first two components of every epistemic field *are* material systems.) The last seven components of the septuple shall be called the *conceptual framework* of the epistemic field under consideration. We shall return to this distinction in Ch. 14, Section 3 when dealing with paradigms and revolutions.

We distinguish two main types of epistemic field: *belief systems* and *fields of inquiry*. In a belief system the components 3rd to 10th hardly evolve or, if they change, they do so exclusively as a result of controversy or of external pressure: witness the ideologies. On the other hand in a field of inquiry every component from the 3rd to the 10th changes in the course of time, and does so mainly as a result of inquiry within the field or within contiguous fields. Besides, every field of belief is self-contained or nearly so: it is neither supported by, nor supports, other epistemic fields: witness pseudoscience. On the other hand every field of inquiry overlaps partially with some other epistemic fields. More on this in Ch. 14.

Belief systems do not always change spontaneously and gradually into research fields. On the other hand some fields of inquiry may gradually degenerate into fields of belief, and this in either of two ways. One is that the given epistemic field was finite to begin with: it had a finite number of problems, all of them soluble in a definitive way, and such that their solution did not pose any further problems. (We call *finite* any such field of inquiry even though we are at a loss to find clear instances.) The other way is by the gradual decay of inquisitiveness among the field workers, i.e. as dogma replaces conjecture and complacent belief takes the place of risky inquiry.

We do not know whether there are any finite and therefore exhaustible fields of belief: all of the known ones seem to be potentially infinite even when they attack finite problems such as censing populations or mapping chromosomes. Indeed, nontrivial knowledge of fact seems always to be spotty and more or less inaccurate—as well as obsolescent because of the extinction of things of some species and the emergence of things of new kinds. No sooner a body of data or of theory is pronounced perfect, than some gap or flaw is discovered in it—and somebody suggests ways of correcting or completing it. Therefore it may be assumed, though perhaps not proved, that, as long as there are workers left in a field of inquiry, the latter will continue to be cultivated—perhaps so intensively and effectively that it eventually turns into a different field. The antecedent of this conditional statement used to be taken for granted since the 17th century, but must now be questioned in view of the increasing virulence of antiscientific ideologies. More on this in Ch. 13, Section 4.

Strangely enough, though inquiry occupies an increasing number of people, and even though we are becoming increasingly familiar with it, there is hardly consensus on its very nature. To some, scientific, technological and humanistic research are arts like poetry or music, for they demand imagination, dedication, and enthusiasm. To others, they are crafts that can be learned and practised by almost anybody. To still others, research is an industry—the knowledge industry. Finally, to many, scientific research and technological development are an immoral racket bent on destroying life. Curiously enough, there is a grain of truth in every one of these opinions. In fact at their best science, technology and the humanities are arts demanding very special callings, talents, and a passionate commitment. At their next best they are nothing but sophisticated crafts, only trickier than the craft of versifying or that of mixing sounds, but still crafts rather than arts. (Still, one draws the difference between mere craftsmanship, however superb, and the inspiration that alone can deliver original art, science, technology, or the humanities.) And it is also true that some research deserves being called an *industry*, namely when it deals competently with pseudoproblems or miniproblems. Finally, it is also true that some scientists, technologists and humanists are crooks—but they are only a few and they are easily found out. In sum, we do not know much about inquiry, beyond that it is interesting and many-faceted, hence worthy of being inquired into.

4.2 *How is Knowledge Possible?*

Is knowledge possible? If so, to what extent and how? The first question may sound facetious but it is a serious epistemological question and one that was not easy to answer before modern science and technology were born, i.e. at a time when most knowledge was mere opinion. Not only skeptics like Sextus Empiricus, Francisco Sánchez, and Unger (1975), but also dogmatists such as Plato and Nietzsche, have denied that we can get to know anything about factual matters. (Mathematics has seldom been questioned.) Their favorite gambit was to exhibit gaps and errors in the bodies of received opinion.

The classical mode of refutation of such radical skepticism concerning the possibility of factual knowledge is still valid. We show knowledge to be possible by exhibiting items of actual knowledge, thus making use of Aristotle's modal principle "If p, then p is possible". (*Trivial examples*: I know the address of my home, the tastes of my children, Maxwell's equations, the empirical formula of water, and the specific function of the heart.) This disposes of radical or *systematic skepticism*. And a consideration of its historical role shows it to be ambivalent for, although it can destroy dogma, it also discourages inquiry.

Methodical skepticism, or *fallibilism*, is a different matter altogether. It is the thesis that, although we can get to know something, such knowledge may be imperfect (neither complete nor entirely true), and is therefore susceptible to criticism and improvement. Fallibilism is a component of *critical realism*, i.e. the epistemology according to which we can represent and understand reality, though not exactly as it is. (More on critical realism in Ch. 15, Section 2.)

Fallibilism is not enough: it must be supplemented with another sobering element, namely a dose of *agnosticism*, i.e. the thesis that there *are* unknowns and unknowables. *Examples*: If the speed of light is indeed the ultimate speed, then we shall never know what is happening at the same time in a distant place; if the big bang hypothesis is true, then we shall never know what went on in the universe before that hypothesized initial explosion occurred; if the brain does not contain detectors and decoders of extremely low energy electromagnetic waves, then we shall never be able to communicate telepathically even at a short distance; if there are no laws of history then we shall never be able to forecast with precision any major social events.

Note the conditional form of our wagers: *If A (is true), then we won't be able to know or do B.* Such conditionalization is a protection against dogmatism. We do not wish to repeat Comte's and Spencer's mistake of drawing long lists of *ignorabilia* that turned out to be subjects of inquiry and even entire scientific fields, such as atomic physics and astrophysics. Comte and Spencer were merely projecting their own philosophical prejudices into the future. Any list of *ignorabilia* that we may care to compile must be based on current knowledge not current ignorance. Thus because we *know* that certain biospecies and human cultures have become extinct, we shall never get to know them adequately. (However, living fossils or remarkably well preserved remains may turn up if we do not give up the search.) Likewise because we *know* that every experiment is subject to accidental errors, we shall never know whether a particular measurement outcome is error-free. (However, we also know that such errors can sometimes be reduced by operating at extremely low temperatures, so there is hope.) In sum, our knowledge is and will always remain incomplete and inaccurate—but there is no knowing beforehand and exactly *what* will remain unknown. We do not know any laws concerning the evolution of knowledge enabling us to make such predictions.

We come now to the third and last of the initial questions, namely 'How is knowledge possible?' This is a classical problem in epistemology and psychology, though one that skeptics (for obvious reasons) and others have refused to formulate. The classical answers to this question are:

(i) *naive realism*: the mind has direct access to things without the intermediary of either the senses or reason—e.g. intuitively or because the brain reflects the world;

(ii) *empiricism*: the mind receives impressions from the senses, so that it knows (only) phenomena (radical empiricism); the mind knows phenomena through perception, and it constructs abstract ideas (logical empiricism);

(iii) *idealism*: the mind grasps self-existing ideas (Plato) or essences (Husserl); there is a pre-established harmony between mind and world (Leibniz); the mind constructs the object, and certain ideas are prerequisites for any experience (Kant).

Although there is something to be said for each of these schools, none of them is fully satisfactory. Naive realism is too optimistic, it has no use for transempirical concepts and mathematical theories, it fails to explain creativity as well as error, and it proposes no cognitive mechanism. Empiricism underrates reason and is therefore impotent to explain abstract ideas such as those of logic, set theory, and abstract algebra. Idealism downgrades experience and does not account for error, let alone for the need to test our hypotheses. Besides, idealism and empiricism ignore the brain and the social context of cognition. Because they place mind out of both nature and society, empiricism and idealism can give no correct explanation of the generation of knowledge. Such reification of mind and its detachment from the world renders knowledge mysterious: in particular it makes science look "miraculously improbable" (Popper, 1972, p. 28). More on all these *isms* in Ch. 15, Section 2, where scientific realism will be defended.

Neuropsychology and the epistemology attached to it approach the problem in a radically different manner. We regard the organ of knowledge (the central nervous system) as a material system placed right in the middle of the natural and social world. As part of the world, the central nervous system can obtain some "inside knowledge" of it. The brain can represent other concrete items by being acted upon by them and generating neuronal configurations similar in some respects to the represented things. In short the brain can know matter because it is material itself. (The mythical immaterial mind, standing off the world, cannot interact with the latter, so it must be assumed to create the world all by itself.) Moreover, because cognition is a process in a concrete thing, it can study itself like any other concrete process. And, because cognition is a biosocial process, it develops and evolves along with other processes occurring in the higher social vertebrates. Thus we can get to know something about knowledge with the help of biology, psychology, social science, and philosophy: the task of epistemology becomes feasible. But at the same time epistemologists must work much harder than heretofore: a priori pronouncements, even if insightful, won't satisfy us any longer.

The conditions that render knowledge possible are exactly those that have resulted in the emergence of animals endowed with a central nervous system. They are extremely special and therefore exceptional material conditions. If the universe were a high density solid block, no knowledge would be possible, not only because life would be impossible, but also because everything would hang together so tightly that knowing a little bit would require knowing the whole. Likewise, if matter were distributed much more thinly, the bonds among things would be so weak that no cells would self-assemble; and, even if they could emerge, they would be nearly isolated from the rest and thus would have no chance to communicate. Also, too short a life span would leave no room for learning: all perception and behavior would be controlled genetically. Besides such physical and biological conditions for learning, think of the social ones: the smartest animals are gregarious. Solitary animals can hardly learn from others. As a Leibnizian would say, ours is the very best world to be known—which does not entail that it was designed so that it could be known.

Finally, how about computers: do they know, and if so, is there any limit to what they can get to know? Most artificial intelligence practitioners claim that computers do know and that in principle there is no limit to what they can know. On the other hand the critics point out a number of limitations—e.g. that a computer can never know any statements lying beyond formal systems (because a computer program is a formal system describable as a Turing machine). Presumably, future computers could overcome or circumvent any such limitations *if* indeed they were capable of knowing *anything*. But they do not know anything if one admits any psychobiological notion of knowledge. The only way out for the computer cultist is to offer a redefinition of knowledge applicable to computers as well as to animals, and one not hinging on the mythical separation between body (or hardware) and mind (or software). So far such redefinition is not forthcoming. So far computers are only cognitive auxiliaries, albeit marvellous ones. They do not initiate inquiry and they do not create any new concepts, let alone hypotheses, theories, methods, or value systems. To hold that computers do know is to indulge in metaphor—like saying that the emergence of every new "generation" of computers constitutes a quantum jump or a mutation.

5. Concluding remarks

Most of us go about inquiring in an empirical way, guided or misguided by intuition and tradition rather than by any explicit epistemological principles. Nonetheless we often manage to join the negative aspect of rationalism (namely its a priorism) with the negative one of empiricism (mainly the collecting of data for their own sake). Little wonder then that most of us harbor heaps of groundless beliefs and act accordingly, i.e. irrationally.

For example, we sometimes believe in the causal or even magical efficacy of the word even in the absence of interlocutors; we seldom reckon with chance coincidences and try to give causal explanations of them; we are prone to believe what we want to believe, and so we tend to disregard or reinterpret unfavorable evidence; we tend to fall for any comprehensive but simplistic and untested doctrine; and we have an excellent memory of our hits but a poor one of our misses. In short, we are often the victims of epistemic accidents.

It may well be impossible to avoid all the epistemic hazards. But with some formal education, some research experience and, above all, with some critical reflection upon our own successes and failures, we are bound to succeed in bringing down the frequency of epistemic accidents. And if we realize also that nearly every epistemology is but a blow-up of some partial truth, we may also avoid espousing it dogmatically, and may continue instead our search for, or updating of, an epistemology free from major blunders and capable of inspiring research instead of misguiding it.

CHAPTER 3

COMMUNICATION

Learning is private but inquiry is social, for it is done in some society or other. The social character of inquiry shows in our dependence upon tradition, in our need to share our questions and answers with others, and in our doing so through language and other, more primitive, modes of communication, such as gestures and calls.

Communication is an essential ingredient of social behavior in all gregarious animals. Through it we share cognitive attitudes and findings, and we influence cognitive processes in, as well as the behavior of, other animals. Through communication we are able to find out part of what goes on in the brains of our conspecifics and, what is more, we can modify such processes and therefore their behavioral outcomes. In this way every cognitive process is a component of some communication network or other—a network that nowadays embraces the whole of mankind. For this reason the lone knowing subject of traditional epistemology is as useless a fiction as the Robinson Crusoe of neoclassical economics.

Stressing the social character of inquiry and communication does not entail ignoring the fact that language is also a powerful thinking tool. As recognized long ago, thinking is to a large extent accompanied by talking to oneself. So, language discharges two functions: a cognitive and a social one. Chomsky (1975) and his followers have stressed the former, studying language exclusively as a mental faculty. On the other hand Wittgenstein (1953) and his followers are mainly interested in language as a mode of social behavior. Actually these points of view are mutually complementary rather than exclusive: speech is a tool of inquiry and learning as well as a mode of behavior and a social glue.

1. SUBJECT AND WORLD

1.1. *Subject–Object Relation*

The pivotal notion of traditional epistemology is the relation between knower (subject) and known (object)—and, in particular, the relation between inquirer and world. Therefore a task of epistemology has always been that of investigating the general problems involving such relation. To be sure, subjective idealism does not face this problem for it conflates subject with object, and so transforms the subject-object relation into a relation between the subject and himself. But in any non-subjectivistic epistemology the distinction between knowing subject and object known (or rather inquired into) is axiomatic. We call it the epistemological *duality postulate*—not to be mistaken for mind-body dualism. (For an examination of realism see Ch. 15, Section 2.2 and Lovejoy (1929).)

The epistemological duality postulate is essential to science, technology, and the humanities. Thus astronomy is the study of celestial objects not of the astronomer's inner life; sociology studies social systems, not the scientist's own "perception" of social facts; technology designs artifacts (physical, chemical, biological or social) instead of studying the daydreams of inventors; and the humanities study the lives and works of thinkers, artists, and men of action, not the scholar's own ideation or behavior. This does not entail that astronomers must ignore the way they perceive celestial objects, or that sociologists must push aside the way they conceive of social systems, and so on. Every specialist must study not only the referents of his field of knowledge but also the ways to study them properly, precisely in order to screen out the distortions ("artifacts") introduced by his viewpoints and techniques. Such self-knowledge, instead of the disregard for the knowing subject, will help attain or improve the objectivity that is supposed to characterize science, technology, and the humanities.

The distinction between subject and object need not, nay ought not to, imply their separation. I.e. the dual*ity* of inquirer and inquired need not imply a dual*ism* wherein the inquirer is divorced from his object of knowledge, and experience and theory are removed from nature and society, to constitute an independent realm. Such epistemological dualism, the partner of the metaphysical dualism of mind and body, is to be rejected if cognitive processes and their externalizations are to be understood as both organic and social processes, and therefore within the reach of scientific research.

A scientific approach to the subject-object relation is then at the same time epistemologically realist and ontologically naturalist (materialist). To be sure, historically not every realist epistemology has been associated with a materialist ontology. Thus Aquinas held consistently a realist epistemology and a hylomorphic and even supernaturalist philosophy of nature. Conversely, it is possible for a materialist not to hold a realist epistemology but to maintain, e.g. that all knowledge consists of conventions and rules designed to facilitate biological and social adaptation. What does force one to adopt at the same time some form of epistemological realism and some form of naturalism is science not logic.

In other words, the subject-object distinction is correct and necessary as long as it is not transformed into belief in an utter separateness of man from his environment, as preached by the monotheistic religions and by idealism. Distinctive we are, separate not. In particular, science and technology are not detached from society but are social activities; and they do not stand across nature but are aspects of man's interaction with his environment. To put it in traditional jargon, the realist epistemology required by scientific, technological and humanistic research should be joined to a materialist ontology asserting the essential oneness, interconnectedness, mutability and lawfulness of all things, inquiring and inquired. (Note that we have written 'things' not 'objects', for there are non-things, such as constructs, that are objects of knowledge. More on this below.)

Another way of putting all this is as follows. The concepts of subject and object may be regarded either as mutually independent (not interdefinable) or as definable. Idealists have tried to define the object in terms of the subject (or of God). In particular, subjective idealists have attempted to define a material thing in terms of the sensations or the perceptions (or in general the ideas) it produces in a subject. Recall Stuart Mill's definition of a thing as a permanent possibility of sensation, and similar definitions by Avenarius, Mach, and at one time Russell and Carnap. Neither of them seems to have realized that any such definition is circular for including the very term to be defined. (Consider: "To be a *thing* is the set of all ideas elicited in act or in potency by the *thing* on some subject".) On the other hand objective idealism can be logically consistent but at the price of postulating a Supreme Being that exudes things as he thinks them up. By so doing the philosopher turns theologian and alienates himself from science and technology.

What about abstract ideas, such as those of triangularity and goodness? Do they not refute materialism and epistemological realism? They do so if construed as self-existing—i.e. provided one postulates immaterialism instead of trying to refute materialism by honest toil. The materialist has no trouble disposing of that opinion. He starts by noting that there is no evidence for the self-existence postulate: that all we know is that some animals are capable of thinking abstract ideas. And he goes on to postulate that such thoughts are brain processes, as assumed by physiological psychology. So, when the subject thinks of triangularity he creates this idea (brain process) instead of apprehending it as in the case of perceiving a mountain. In other words, if the object is an idea then it is part of the subject. In this case strictly speaking—ontologically—there *is* no subject–object relation. In this case there is only a cognitive activity of the subject—in particular her feigning that the "objects" of her cogitations are self-existing.

This has important consequences. One is that not every object of knowledge is a material object existing independently of every possible knowing subject: some objects are parts (actual or potential) of some subjects. Another consequence is that, paradoxically, the knowledge of logic, mathematics or philosophy is, strictly speaking, a form of self-knowledge, because original logical, mathematical and philosophical objects are "produced" by the self instead of being self-existing. However, they have the appearance of objectivity and publicity because we have purposively stripped them of any idiosyncratic features. In other words, abstract ideas are formed by abstracting from personal characteristics and circumstances: they do not bear any birth marks and so are universal, i.e. thinkable in principle by anyone (Bunge, 1981a). In contrast, my knowledge of my present mood is only mine (unless I care to disclose it to somebody else) and, for being mine, it is not universal. Which brings us to the problem of the knowledge of mind.

There are multiple accesses to the brain processes of the mental kind, from physiological studies of brain activity and investigations of the behavioral manifestation of such activity, to external documents such as things fashioned or destroyed by the subject, to introspection or its verbal expression. The latter has been discredited by behaviorist psychology but it still is, and will always remain, an indispensable source of psychological knowledge, particularly when complemented and controlled by the other sources. Only, we must not say that 'the mind knows itself by introspection, i.e. by contemplating itself'. A linguistic behaviorist might say that this is bad grammar, because contemplation is an irreflexive relation: a can contemplate b only if a and b are different. We say that view is bad biology, for it reifies functions (activities). What can be said instead is that, in introspection, one part of the brain inquires into what another part of the same brain is doing, or what one subsystem of the brain did a moment ago: introspection, like self-consciousness, is just the self-monitoring of the brain. Moreover introspection, though necessary to understand the mind, is insufficient, for it tells us only *that* we are having such and such subjective experience, not *how*, let alone *why*, we are having it.

1.2. *Inter-Subject Relation*

Because he can communicate in some form or other, every human being can teach others and learn from others—sometimes long departed ones, thanks to cultural externalizations such as written documents, drawings and tapes. This does not mean that cognitive attitudes, processes and their outcomes can be transferred the way goods change hands: learning something from someone else consists in going through certain similar brain processes that are not transferable and that most of the time cannot occur unless previous learning processes have taken place.

Communication is a kind of social behavior and it occurs in communication systems, be they relatively simple ones like the mother-infant system, or complex ones like an international television network. Every communication system is composed of three main components: sender, channel, and receiver. The system is active every time the sender emits a message that gets transmitted (usually with some distortion) through the channel, and the receiver succeeds in decoding it. Any of these components or even all three can be nonliving, but there is no communication system unless at least the sender is controlled, directly or indirectly, by some animal. (The receiver can be a robot. The sender too can be one, but it will function on behalf of some living being.)

Of all communication systems the most effective ones are those where transmitters are also receivers, i.e. where there is a two-way communication channel. The traditional classroom, with its lecturer and his passive pupils, was a one-way communication system. The modern class, where feedback is encouraged, is a two-way communication system that facilitates error correction as well as learning on the part of the teacher. Actually every society is, among other things, a communication system, for communication is a part of social behavior and therefore a member of the social structure (collection of social relations.) It is not only that cognition is influenced by social relations: every communication act, from asking a question to answering (or refusing to answer) it, from exchanging gossip to trading data, from listening to the news to selling a newspaper, is a social transaction that contributes to either social cohesion or social disintegration, to the growth or the decline of knowledge.

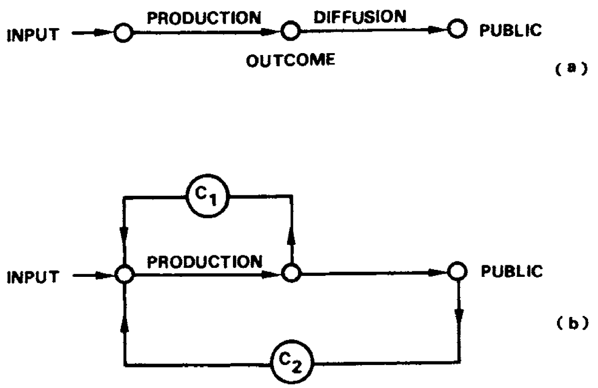


Fig. 3.1. Two views of the process of production and diffusion of knowledge. (*a*) Traditional views: open loop (no control). (*b*) Contemporary view: closed loops double control system: the subject corrects her own inputs (data, problems, etc.) in the light of her output, and society checks the whole.

From the point of view of systems theory, a communication system is a control system where the controls check the outcome of the process. Since there is the control by the knowing subject and by the society in which she is embedded, the system has two closed loops: see Figure 3.1. Without such control, irrelevance and error would go uncorrected, and society would neither encourage nor discourage any line of inquiry—for better or worse. Needless to say, the controls are in the hands (or rather brains) of all concerned, senders and receivers, producers and consumers of knowledge. This model contrasts with the open loop (no control) model of both empiricism and idealism: Table 3.1. The open loop model includes no mechanism, whether individual or social, for error correction, other than

TABLE 3.1
Three models of the production and diffusion of knowledge

	Empiricism	*Idealism*	*Realism*
Inputs	Data	Background knowledge	Background knowledge, data and problem
Transfer	Induction	Hypothesizing	Creative process
Output	Conclusions	New knowledge	New knowledge and problem
Control 1	——	——	Error correction
Control 2	——	——	Social checking

starting all over again, i.e. modifying the transfer function so as to obtain a different system. No such system change is necessary in the case of a closed loop control system.

We have avoided the term 'information' for several reasons. First, it is ambiguous, as it stands for both the act or process of conveying a message and for the content of the latter. Second, we communicate more than information or knowledge: we also transmit questions and commands, and we convey epistemic attitudes such as doubting, asserting, and denying. Third, since the inception of information theory (Shannon and Weaver, 1949), the term 'information' has acquired a technical signification that does not coincide with the ordinary language term, let alone with the traditional word 'knowledge'. Indeed information theory is concerned with the process of message transmission, not with what gets transmitted, which may be an equation or a nonsense string, a question or an order. The theory—like the communication systems it deals with, such as the mail and telephone services—is not concerned with the content of the messages the transmission of which it studies. Therefore information theory, though sometimes regarded as the basis of epistemology and semantics, is hardly relevant to them.

We shall have more to say on communication in Section 3 after having taken a look at the wider issue of the social matrix of the production and diffusion of knowledge.

2. SOCIAL MATRIX

2.1. *Cognitive Habitat and Niche*

Every animal has a habitat or region where it can live, as well as a niche in the ecological network. And every animal capable of learning has, within its total habitat, a *cognitive habitat* or region favorable to its cognitive activities. Thus at present the whole Earth and the Moon compose the human habitat, which may soon expand to encompass the entire solar system. But the Moon is not a good place to study anthropology, and Baffin island is not suitable for studying tropical zoology. Every human habitat is suitable, or unsuitable, to some degree or other, to learn about something.

In addition to a cognitive habitat, every animal capable of learning has a *cognitive niche* or set of possible cognitive states—or collection of "things" it can get to know. (A cognitive niche can be construed as a box in the corresponding cognitive space, or state space of the cognitive subsystem of the nervous system. Every cognitive niche is included in the total niche of the animal.) Animals belonging to different species can have very different cognitive niches, which sometimes hardly overlap. Think of man, mole, and louse.

The cognitive habitat of an animal capable of learning includes its parents and all the conspecifics with which it is in touch. In the case of social animals such contacts are not sporadic and random but permanent and regular. In this case every animal can learn from every other member of its community, and can teach some other animals no sooner it has learned something as well as how to communicate it. Learning is a social enterprise. We do not start from scratch but from the findings (and limitations and failures) of our teachers, and we learn from one another. Tradition and the flow of information are particularly strong among humans, and they are getting ever stronger, particularly with the establishment of communication links and the invention of cognitive records such as drawings and writings. However, this social aspect of learning should not be exaggerated: we learn *in* society but it is the individual, not society, that does the learning, for society is brainless.

The union of all the individual cognitive niches may be called the *public cognitive space*, or the total cognitive space of society (or mankind). One advantage of an (ideal) free society would be that each member of it would have access to any region of this public cognitive space, i.e. to the cognitive abilities and achievements of his fellow men. This possibility does not exist in societies divided into groups (e.g. priestly, professional or military castes, political parties or corporations) that exert a monopoly on cognitive activities of certain types. Where the public cognitive space is not partitioned into water-tight cells, an individual can learn much and quickly.

Man does not evolve certain cognitive abilities except in certain societies. For example, hypothetical reasoning, or the ability to reason in terms of *ifs*, is the principal novelty in the cognitive development of the child reaching adolescence in urban communities. This stage is seldom reached by people living in the countryside or in backward villages, let alone in preliterate societies. Here the subject is incapable of adopting a detached attitude and of contemplating situations or views other than his own, or of taking an interest in questions that go far beyond his immediate field of interest (Piaget, 1972; Luria, 1976). Schooling, even in the village, has dramatic effects, not only on cognitive attitudes but also on affective ones such as punctuality, discipline, health, nutrition, family size, and sensitivity to public issues—all of which have in turn a noticeable effect on economic growth (Lewis, 1955; Denison, 1962; Becker, 1964). In sum, human cognition is social through and through. Consequently psychology cannot study it properly unless it joins forces with sociology.

Take the social inhibitions, as well as the lack of social stimuli, to inquiries of certain types. A random sample must suffice. (*a*) The monopoly on literacy and inquiry, which was universal until recently, is a major obstacle to inquiry. (*b*) The extreme division of labor between manual and intellectual workers, and the low social status of the former, has been a major obstacle to scientific and technological innovation: craftsmen had no incentive to invent or to adopt new ways of increasing the quantity or the quality of the product, and intellectuals to engage in experimental inquiries. (In some nations, such as the U.S.A. and the U.K., scientists and technologists are not supposed to do their own machining.) (*c*) The prevailing value system may place little if any value on basic knowledge—this being why China, India and Islam did not contribute significantly to the building of modern science (Roche, 1976). (*d*) The prevailing ideology may discourage or even ban entire lines of research—e.g. of evolutionary biology in a theocratic state, and of social science in a totalitarian one. (*e*) A dogmatic school system will inhibit inquisitiveness and a critical attitude—whereas an excessively unstructured school will fail to teach intellectual discipline. (*f*) The preparation for, and waging of, war may accelerate the pace of destructive technology (drawing on a fund of basic knowledge). But it interrupts basic research by diverting human and economic resources—and by killing budding investigators in the battlefield. (*g*) The censorship exerted by organized dogmatism (church, party, clique, school) protects error from criticism and blocks the diffusion of new approaches and new results incompatible with dogma. Whatever does not fit with the latter remains unnoticed or, if noticed, is suitably "interpreted" (Fleck, 1935). (*h*) Intellectual fashion spurs research into the fad concerned but, by the same token, it diverts attention from other fields. (*i*) The current method of funding research encourages safe investigations expanding existing fields rather than opening up new ones. Indeed peer review tends to guarantee orthodoxy and suppress originality, encouraging researchers to swim in the main stream. (*j*) The requirement to remain within a given specialty makes for narrowness and blocks the diffusion of knowledge across cognitive fields. Trespassing should be rewarded not penalized, for it often gives rise to new approaches and occasionally even to new fields—as was the case with operations research, computer science, molecular biology, and physiological psychology. (*k*) Society is not always quick to adopt improved ideas, procedures, or things. It will accept a novelty only if it is quite obvious that it fills a vacuum. Even so the resistance of people who have vested interests in the old must be overcome or bypassed: remember that "One person's innovation is ordinarily another's destruction" (Brewer, 1980). See Figure 3.2.

A few centuries of relentless growth of knowledge has made us forget that this is not a law but a recent trend, and that the right to learn, like any other right, is not a birthright but must be conquered or defended. In some societies most of the people are prevented from learning certain skills—if not by force at least because learning demands some leisure. Only a few societies recognize the right to know and make it effectively possible for everyone to learn. In most societies the individual is denied the right to learn about the major decisions that may affect him or her, and a fortiori to participate in the taking and implementing of such decisions. Every society has economic, political and cultural mechanisms for controlling the production and flow of knowledge: no modern society can afford to leave cognition totally in the hands of private initiative. (More in Ch. 13, Section 4.)

Having recognized the need to study the social matrix of cognition—and therefore the legitimacy of the sociology of knowledge—we should sound a warning against epistemic *sociologism*. This is the view that every cognitive activity and every bit of knowledge should be studied and evaluated *only* in

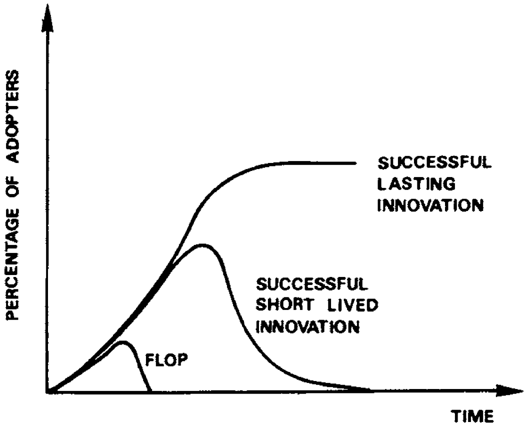


Fig. 3.2. The adoption of technological innovations.

the light of society and its history. Sociologism tells us that in the final analysis there is but one science: the science of man (Comte); that consciousness is determined (not just influenced) by socioeconomic conditions (Marx); that ideas are not so much in individual brains as in society—e.g. that Newton's physics was but a product of the first Industrial Revolution (Hessen, 1931); that the facts studied by science are not objective but are manufactured by the scientific community (Fleck, 1935); that truth is ultimately what the community of scientists decides to accept (Kuhn, 1962); that what distinguishes science from other cultural pursuits is its social condition (Ziman, 1979); and that an individual of genius is not one whose work is unique but, on the contrary, one "whose work in the end would be eventually rediscovered" (Merton, 1973)—so that a scientist of genius is "functionally equivalent" to a bunch of mediocre researchers. (By the way, Kuhn (1978) seems to have renounced sociologism to revert to internalism.)

Sociologism is wrong for the following reasons. First, all cognitive processes occur in some brain or other. Even a public utterance, or a formula written in a book, do not become cognitive items (i.e. processes in some plastic neural system) until somebody takes cognizance of it. Second, powerful new ideas and shattering discoveries are made only by exceptionally talented individuals: without these most of us would have no chance of participating, whether as beneficiaries or as victims, in any scientific or technological revolution. Third, although acceptance or rejection by a community are important to the fate of any new idea, popularity is not an indicator of truth anymore than impopularity an indicator of falsity: truth is a semantical and methodological category, not a sociological one. Fourth, although modern science and technology prosper only under exceptional social conditions, these come in a great variety, from capitalism to socialism, and from industrialization to semidevelopment. Moreover a mark of modern natural science is that its results hardly bear the seal of social circumstances, which on the other hand continue to exert deep influence on social science research. (For example, until the Great Depression economists were hardly concerned with unemployment, which since Keynes' *General Theory* (1936) has remained a crucial macroeconomic variable.) In short, society stimulates or inhibits inquiry in all fields, but does not do the inquiry because it has no brain. We do not have to choose between the isolated knowing subject of classical epistemology and psychology, and the all-powerful society imbued with a spirit of its own: we can choose to study the more modest but real inquiring subject-in-society. See Figure 3.3.

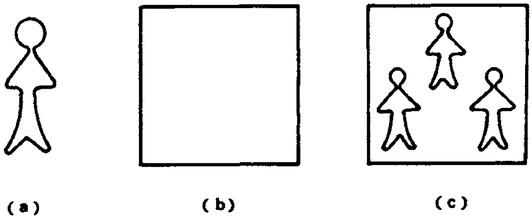


Fig. 3.3. Three views of cognition. (*a*) Traditional epistemology: the isolated knowing subject. (*b*) Sociologism: the brainless social framework. (*c*) Social neuropsychology: the community of learners embedded in a wider social matrix.

2.2. *Learning Community*

Every human being is a member of some *inquiring system* or *learning community*—i.e. a system of persons who learn by themselves as well as from one another by exploring and thinking, imitating and teaching, questioning and criticizing. And every learning community, whether advanced or primitive, has one or more *learning professionals*—medicine men or scribes, craftsmen or engineers, scientists or philosophers, teachers or journalists. Although these professionals may discharge other functions as well, their specific function is that of acquiring, utilizing or diffusing knowledge. These learning professionals form a subsystem of the cultural system that composes every community regardless of its level of development. Each of them holds information relations with other members of the community: some of them direct or interpersonal, others indirect, i.e. via couriers, correspondence, or publications. The potential clientele of these learning professionals is society at large.

In a modern society every learning community is divided into four main subsystems: the scientific, technological, humanistic, and artistic communities. Each learning professional is likely to be a member of several subsystems of the total learning community: e.g. a laboratory or some other unit, a university (or firm or government) department, a chapter or a professional society, an "invisible college" formed by knowledge consumers and evaluators, etc. In these conditions no learning professional is free from influences or completely devoid of cultural power over others. Typically, the crackpot is marginal: he is deprived of the benefits of stimulation and feedback. ("It is astonishing what foolish things one can temporarily believe if one thinks too long alone": Keynes, 1936, p. xxiii.)

The active member of a learning community shares in upkeeping certain inquiry traditions, updating them, and introducing new ones.

The organization of society as a whole, as well as that of each of its learning communities, is no less important for the conduct of inquiry than native intelligence, motivation, and industry. A mediocre intelligence can make substantial contributions to knowledge provided she is a component of a well organized learning community, whereas a genius is likely to go to waste in a cultural desert. The quick pace of innovation in science, technology, the humanities and the arts since the dawn of the modern era can be explained in terms of the emergence of a new type of society, more receptive to novelty and more mobile than the feudal society. The earlier explosion of knowledge that accompanied the birth of civilization about 5,000 years ago, and the no less dramatic explosion that accompanied the emergence of agriculture about 15,000 years ago, are further examples of the connection between modes of inquiry and social structure. Also, the revolution in technology and art that occurred about 30,000 years ago is likely to have accompanied a social reorganization which may have been prompted by a steep increase in population density, which in turn must have strenghtened and quickened the information flow.

Every specialized learning community, such as the community of biochemists or that of philosophers, is a social system. As such it can be represented by the triple *composition–environment–structure* (Vol. 4, Ch. 5). The *composition* of the community that cultivates a given field of science, technology, art, or the humanities, is of course the collection of all the workers in that field plus the support personnel (technicians, clerks, librarians, etc.) and their work tools. The *environment* of the community is the collection of all the items in the natural and social environment of it, including the subject matter in case this is material (e.g. the biosphere in the case of biologists, the whole universe in the case of cosmologists). And the *structure* of the community is the collection of relations of inquiry, teaching, receiving or conveying information, managing, etc., that give the learning community whatever cohesion and efficiency it may have. (We say that these activities are *relations* because this is what they are from a logical point of view. Thus inquiry is a ternary relation, for the standard form of a statement including the concept of inquiry is this: A inquires into B by means of C. Likewise teaching: A teaches B to do C.)

The study of learning communities is of course a task for social scientists, but their study as sociosystems has only begun. However, in order to understand any such community one must be able to understand what its members do: what problems they face, how they handle them, what kinds of hypotheses, data and methods they use, and so on. Unfortunately all too often the anthropologist or sociologist studying learning communities ignores this caveat and limits his study to observing the motions through which his subjects go—which is compatible with behaviorism but far off the mark of a scientific study of cognitive activities. Not surprisingly, the results of such studies are trivial or even grotesque. Ignorance is no substitute for inquiring about inquiry.

Every normal social system is propelled by cooperation as well as competition. The learning communities are no exception. Cooperation includes in this case help in transmitting skills and information, formulating or reformulating problems, devising hypotheses or methods, offering constructive advice or criticism and, in general, sharing knowledge. Competition includes making destructive criticism, proposing rival theories, methods or data, and fighting over the support of third parties, such as assistants, colleagues, and funding agencies. Cooperation gives individual workers the necessary support, and competition keeps them on their toes; competition helps find out error, and cooperation corrects it. Where cooperation falters, the system disintegrates; where competition disappears, the learning system may degenerate into a belief system.

In an increasing number of fields inquiry is conducted by *research teams* animated by an *esprit de corps* and where the division of labor, sometimes excessive, is compensated for by efficient organization. The research team, particularly conspicuous in natural science and technology since World War I—but still exceptional in the humanities—was a natural offspring of the scientific enterprise led by an autocratic leader. The research team is often democratic and can handle not only cross-disciplinary research but also unidisciplinary inquiry involving a number of skills unlikely to be mastered by a single individual—e.g. theorizing, computing, designing experiments, measuring, and performing statistical analyses. The publication summarizing the results of the inquiry is usually signed by the principal investigators, who acknowledge the help of their assistants, or by all the members of the team even if they number twenty or thirty individuals. Team research is far more common in experimental and applied science than in theoretical science. It is most successful when research can be planned in detail—which is possible only when known methods are employed to attain specific goals in a planned manner, which is not the case of the exploration of the unknown (Abelson, 1965).

The *research factory* is a different kind of inquiring system. It is usually set up by a captain of industry or a government official who controls resources, hires and fires, establishes the goals, and sketches the research or development project. Edison's Menlo Park, the Manhattan Project and the Apollo Mission are good examples of successful research factories. They were successful because their members agreed on a corpus of scientific or technological knowledge, on the proper methods to test new ideas and new artifacts, and on the ultimate goal. When such fundamental agreement is missing, mediocrity or paralysis is bound to set in. Although most research factories do not congregate large numbers of star researchers, they do ask for the advice of outside consultants capable of solving some of the toughest problems encountered by the team. Because research factories are mission-oriented, they discard any new problems and new ideas unlikely to be useful to attaining their goal. (Those who shun originality seldom get it.) Moreover the investigators employed by industry are not free to publish their findings, for the latter belong to their employers, who are anxious to patent and exploit discoveries, not to make them available to society. So, research factories curtail both the freedom of inquiry and the freedom of communication.

A third type of inquiring system is the *committee*. Committees may produce, analyze and evaluate research policies and projects, but of course they do not conduct original inquiries. Because most committees are watchdogs rather than explorers, they tend to be conservative. Galileo was condemned by a committee, and Einstein by political bureaus; likewise Marx and Darwin are still being condemned by churches and political groups. Since most people are afraid of novelty—either because they do not understand it or because they have vested interests in the old—the best result a committee can come up with when faced with radical novelty is an ineffectual proposal—such as referral to a subcommittee. Democracy has yet to solve the great problem of combining expert judgment with rule by committee. The key is agreement on both goals and means—a tacit agreement so much in evidence in the thinking and action that went into the selection of the horse and the camel. (It is false that a camel is a horse designed by a committee. What is true is that a committee of inexperts looks as though it were designed by a horse or a camel.)

The local, national, regional and international learning communities are the trustees of knowledge and they provide the mechanism for the steady production and diffusion of it. At the time of Galileo and Descartes the scientific community, then concentrated in half a dozen countries, counted about 200 scholars, mostly amateurs. (Armand Beaulieu, personal communication.) Today the scientific and technological community is spread over the entire world and is at least two million strong, not counting the support personnel. These huge numbers facilitate cooperation and intensify competition. The outcome is not only greater productivity but also higher and higher empirical standards. The ultimate test of a scientific, technological, humanistic or artistic innovation is its achieving international recognition.

The existence of learning communities explains simultaneous discovery and invention, which are otherwise mysterious. Popular explanations for such events are that certain ideas are "in the air" (objective idealism), or that they are necessitated by society, particularly the economy (vulgar materialism). The truth is that coinventors or codiscoverers share (*a*) certain brain properties, (*b*) certain work habits and values, (*c*) a tradition, and (*d*) a set of problems. If half a dozen talented people share a body of knowledge and values, and tackle a given problem at about the same time, it is not unlikely that a couple of them will come up with the same solution at about the same time. Cultural parallelism is no more mysterious than genetic parallelism.

We wind up this section with a couple of remarks. The first concerns group learning: it is sometimes held that not only individuals but also human groups can learn (Botkin *et al.*, 1979.) Literally taken this is impossible, for groups do not have a brain of their own. What is true is that (*a*) individuals in a group can learn from one another, and they can learn to think, feel and act so as to optimize the efficiency and the cohesion of the group, as a result of which (*b*) the group evolves.

Our second remark concerns the ontological status of inquiring systems. At first sight they are not concrete or material entities because they lack most of the properties characterizing physical or chemical systems, and because they make use of constructs. However, (*a*) any system composed of material entities, such as human beings, can be said to be material itself, and (*b*) inquiring systems are not composed of persons and constructs on a par but only of persons endowed with brains capable of conceiving constructs. The material/conceptual dichotomy is methodological, not ontological: it is only a convenient fiction to abstract ideas from the brains capable of thinking them up. (Cf. Bunge, 1981a.)

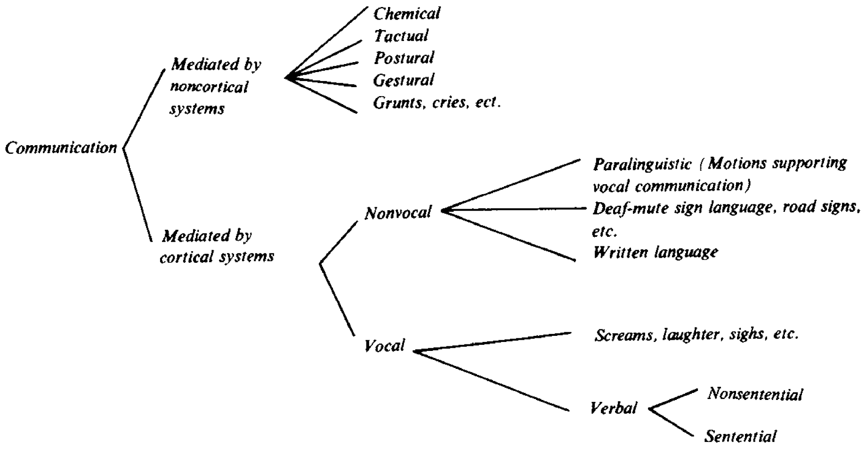
3. Communication

3.1. *Communication System*

All animals endowed with a central nervous system, from ant to primate, seem to be able to communicate, i.e. to send and receive signals carrying some information. (Cf. Sebeok, 1977.) Communication is particularly developed in social animals, where it discharges at least the following functions (Lindauer, 1971): (*a*) to inform other animals that they belong to the same species or even the same society; (*b*) to inform others what their position in the social organization is; (*c*) to inform others in what state the individual is (e.g. whether it is ready to court, fight, or communicate where food can be found); and (*d*) to give or receive instructions concerning tasks of social utility (e.g. to bring in food or take out refuse). Communication, then, is a powerful tool of biological adaptation and social organization. However, here we are mainly interested in communication as a means to acquire, elaborate, and share knowledge.

The key concepts involved in any comprehensive discussion of communication are those of signal, message, significance, understanding, and communication. Definitions of these concepts in psychobiological terms are proposed elsewhere. (*Treatise*, Vol. 4, Ch. 4, Def. 4.55, and Bunge, 1980a, Ch. 9, Def. 9.8.) Here we shall concentrate on communication systems and, in particular, languages. A *communication system* is a concrete (material) system composed of animals of the same or different species, as well as nonliving things, in some environment (from beehive to the universe), and the structure of which includes signals of one or more kinds. Signals are processes executed or controlled by some animals and capable of carrying messages to other animals able to decipher them. These signals must be physical or chemical processes: the mythical psychical waves do not work because they do not exist. And the communicating animals must be endowed with suitable sensors: there is no extrasensory communication.

There are a great many kinds of signal employed in animal communication:


Communication
Mediated by noncortical systems
Chemical
Tactual
Postural
Gestural
Grunts, cries, ect.
Mediated by cortical systems
Nonvocal
Paralinguistic (Motions supporting vocal communication)
Deaf-mute sign language, road signs, etc.
Written language
Vocal
Screams, laughter, sighs, etc.
Verbal
Nonsentential
Sentential

We are more interested in signals systems than in occasional stray signals such as grimaces, screams, or grunts. A *system of signals*, or *communication method*, is a (natural or artificial) collection of interrelated signals, involved in some communication systems, and allowing an animal to compose and convey messages addressed to other animals (primarily but not exclusively conspecific). A system of signals can be either inherited or learned, it may or may not contain a set of basic signals (e.g. words), and it may or may not be composed of an indefinite number of signals. Human speech and the American Sign Language, drawing and sculpting, pictographic and phonetic writing, musical notation and chemical notational conventions, mathematical and computer languages, are artificial systems unique in the animal kingdom.

Take for example the American Sign Language (ASL). Its expressive power is said to be as great as that of ordinary spoken language, though its rate of production is half the rate of the latter (Bellugi and Klima, 1979). Another example of a language is chemical nomenclature. The symbols for the 100 or so chemical elements, together with a set of numerals, suffice to form (by concatenation) the stoichiometric formula of any chemical compound. But in this case the syntax is neither conventional nor a product of a haphazard historical development: it summarizes chemical laws.

A *language* is a system of signals that are invented or learned rather than inherited. Moreover every signal in a language is either a simple signal or a complex one formed by concatenating two or more simple signals according to definite (syntactical) rules. There is no limit to the number of signals (hence messages) that can be formed in a language out of simple signals (e.g. words). (Mathematically speaking, a language is embedded in a free semigroup on a set of basic signals. Caution 1: The standard description of a language as a free semigroup is incorrect because in actual fact we can form only finite collections of strings of symbols of finite length. Caution 2: Some English words do not combine associatively. Thus '(small car) buyer' is not the same as 'small (car buyer)'.)

Many systems of signals, such as road signs, the human "body language", and the signals used by social bees, are not languages proper. Neither of them has a syntax, and the last two are inherited or instinctive rather than learned. Nevertheless some students of animal communication regard the systems of signals used by bees and ants as sets of conventions. Thus they state that taking the sun as a reference point, and "reading" the number of waggles or sound bursts as measuring the distance to the food source, are so many conventions. (See, e.g. Brines and Gould

1979.) But surely this is a misuse of the term 'convention', for it is admitted that such "rules" are wired in the insect's brain.

Human languages, on the other hand, are multiple, conventional, learned, and changeable. They are conventional because there is no law of nature commanding to name things in any one way, or placing noun and adjective in one order rather than another, and so on. (Caution: Although natural languages are conventional they are not the product of deliberate agreements but of cultural evolution.) Consequently any given linguistic convention can be replaced with another or even dropped altogether. And new conventions can be introduced provided they satisfy certain conditions ensuring over-all consistency as well as respecting what the French call *le génie de la langue*—whatever this may turn out to be.

So far as we know only humans have invented conventions and, in particular, linguistic conventions. However, there are indications that anthropoid apes can learn some of our linguistic conventions. A number of workers (Gardner and Gardner, 1969; Premack, 1971; Savage-Rumbaugh *et al.*, 1978; Patterson, 1978) contend that chimpanzees such as Sarah, Lana and Washoe, and gorillas such as Koko, have learned certain human languages such as ASL. They would master them to the point of communicating with their trainers as well as with some conspecifics, and to convey new thoughts not taught them, as well as to talk to themselves. The critics claim that these animals just respond to their teacher's previous utterances, that they have learned to use certain signals by sheer conditioning, and that the system of signals they use is devoid of syntax (Terrace *et al.*, 1979). The whole thing, according to them, would be a delusion, like the Clever Hans phenomenon, resulting from careless methods and anthropomorphic interpretation (Umiker-Sebeok and Sebeok, 1981).

The linguists who have become involved in the controversy—and there are not many—tend to disbelieve the thesis that beings other than man can make use of language. In particular, the transformational linguists pose the problem as follows. The fountain of a language is its grammar, which is composed of syntactical, semantical, and phonological rules for the production and understanding of correct sentences of the language. If an animal is capable of forming new correct sentences—new, that is, to the animal, and which is has not learned by rote—then it can be said to have mastered (to some degree) the grammar and thus the language in question. If on the other hand the animal does not form consistently any new and correct sentences, this is a clear indication that it does not possess a set of rules (a grammar) to generate the sentences of the language. Now, the empirical evidence for "ape language" is weak: it refers to a handful of apes and there is no hard evidence that they do produce sentences really new to them, rather than repeating sentences formed by their trainers. Some workers hold a stronger view, namely that (*a*) there is no favorable evidence whatever, and (*b*) there *can* never be any such evidence, for language is built into the human mind—perhaps to the point of not being learnable at all. Needless to say, there is no evidence for these countertheses either.

It would seem that the investigation into "ape language" should start all over again and be conducted in a more rigorous and thorough manner. In particular the following points should be taken into account. First, a more adequate definition of a language (in particular a human language) should be proposed. (I do not mean a more exact definition, for the one offered by the mathematical linguists, namely in terms of the concept of a free semigroup, is exact. What I mean is a definition fitting any ordinary language.) Second, not only apes but also other animals, in particular "talking birds" and songbirds, should be investigated. Third, the performance of apes and other animals should be compared to that of children for, after all, we hardly know how a child acquires language (Bindra, 1981). It may well turn out that a child's speech, too, is cued by his interlocutor, who may in turn be a victim of the Clever Hans phenomenon. Fourth, the following hypothesis is worth being investigated. When an animal learns to name things or events, it may proceed as follows. In a first stage the object and the symbol (word, chip, figure, etc.) elicit each a process in some neural system. Learning to name the object by the symbol consists in connecting the two psychons or in establishing a third psychon for both. And this is noting but conditioning, that *bête noire* of mentalism: Figure 3.4. As for new sentences, they may result from the sequential activation of psychons for learned words, or from the spontaneous activity of psychons of new types. Since apes and other animals have given ample evidence of inventiveness in other fields, it should come as no surprise that they can learn sentences and even form new ones. But this hypothesis has yet to be confirmed by carefully controlled experiment.

Why has the problem of "ape language" stirred such spirited controversy? There are several reasons for this. One is that the problem is tough, the evidence inconclusive, and our knowledge of the way humans learn to speak is not far superior to our knowledge of the way Sarah, Washoe, Lana, Koko and a few other cousins have learned to communicate using man-made signals. Another reason is the novelty of the

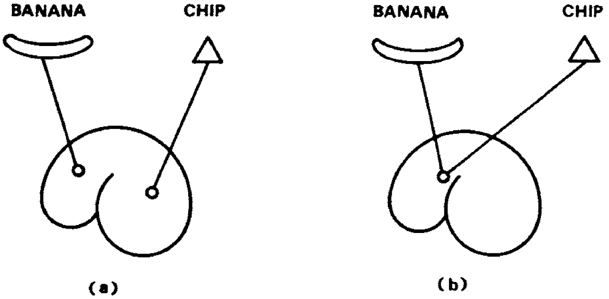


Fig. 3.4. Learning to name an object. In a first stage object and symbol activate different psychons. In a second stage a single psychon is activated.

thesis that animals other than men can talk, particularly in view of the traditional definition of man as "the talking animal". If true, the thesis would have profound implications for psychology (particularly psycholinguistics), anthropology (particularly palaeoanthropology and prehistory), and evolutionary biology (in particular that of primates). A third reason is that the thesis has ideological overtones. Indeed if the thesis is false, then it becomes easier to defend the view that the human mind has nothing to do with subhuman brains: there would be an unsurpassable chasm between man and beast, which evolutionary biology could not hope to explain. Thus the refutation of the hypothesis of "ape language" (or rather that apes can be taught some human languages and the mode of thought that goes with them) would bring comfort to the various religions and spiritualistic philosophies that hold .the dogma of the spiritual (and possibly partly supernatural) nature of man. Actually if the "ape language" theses were refuted the materialist need not concede defeat: he could still hold that language and thought are brain functions that have evolved biologically and socially from humbler neurophysiological functions.

Unfortunately we know next to nothing about the origin of language and, because any prehistoric and palaeoanthropological data we may get hold of will at best serve as circumstantial evidence, we may never be able to formulate anything beyond clever conjectures. (Cf. Lieberman, 1975; Harnad *et al.*, 1976.) However, since language is *inter alia* a means of communication, it is reasonable to conjecture that every major leap in the evolution that led from our remote ancestors, a few million years ago, to *Homo sapiens sapiens*, was accompanied by a major advance in communication. It seems also reasonable to hypothesize that the language of primitive man was quite different from ours: for one thing it may have been just a collection of words, and thus as agrammatical as baby talk. There is no reason to believe that it satisfied the definition of a language fashioned to fit modern languages. Unfortunately most linguists seem reluctant to accept any alternative definition, and thus they close the door to the very investigation of the evolution of language.

To return to the problem of language acquisition. Thanks to the work of developmental psychologists—Piaget in the first place—and neurolinguists, we are learning a bit about the development and loss of linguistic abilities. For example, it seems that language and substantive knowledge are functions of different brain systems. This is suggested by familiar experiences such as forgetting what a name stands for, and what the name of a known concept is. It is also suggested by the observation that a (semantic) aphasic knows that a cup is to drink from even if he is unable to name it. Moreover the surgeon can cut some of the neural links between speech and thought, as shown by commissurotomy patients, who are unable to say some of what they think. (A right-handed person who has had his corpus callosum cut cannot react verbally to verbal stimuli projected to his right hemisphere, but he may produce nonverbal responses indicating that he has understood the message conveyed by the stimulus.)

Furthermore it is possible that syntax and semantics are in charge of different psychons, as suggested by the finding that some neurological (and philosophical and sociological) patients produce gibberish sentences that are grammatically well formed. Also, reading and writing would seem to be controlled by different brain centers. Indeed severe dyslexia patients can write sentences which they are unable to read. Neurophysiological examination points to lesions in the angular gyrus (and usually also in the splenium) as causing dyslexia without dysgraphia. And lesions in the supramarginal gyrus have been shown to cause dyslexia with dysgraphia.

Further confirmation of the hypothesis that language and knowledge are functions of different brain systems is provided by the well known fact that illiterates and deaf-mutes can learn to play chess, paint, draw blueprints, and read maps, neither of which involves verbalized thinking. (Chess thinking is imagining moves, and paint thinking is imagining shapes and colors.) The very existence of nonverbal or nonsymbolic thought refutes the view, rather popular among philosophers from Plato to Quine, that all thought is symbolic or verbal, so that (methodological consequence) the study of language exhausts that of thought. Once more, empirical research has killed philosophical speculation.

Thought and language are then different. But they are coupled: learning to speak goes hand in hand with learning to perceive and to think, and each aids the other. Besides, using a language is a cognitive ability, so that knowing a language is a kind of knowledge, though not one of subject matter. Also, the more substantive knowledge we want to carry, the more linguistic container we need. In general all cognitive abilities—and language is one of them—are linked, and linguistic dexterity helps acquire and perfect other cognitive skills. For example, bilingual children all over the world do better at problem solving and "divergent thinking" (inventiveness) that unilingual ones (Lambert, 1981).

Our knowledge about language acquisition (and unlearning) is insufficient to give definite answers to certain questions that have given rise to a great deal of controversy in recent times. (The temperature of a debate is an ambiguous indicator: it may indicate profound ignorance of the subject, ideological implications, or both.) One such question is whether there is a special "capacity" for learning language—assuming we know what a "capacity" is. Whereas empiricists claim that language is learned like anything else, innatists hold that humans are born with a special capacity for learning languages—and some of them go as far as claiming that we are born knowing the main features of language. These conflicting theses are difficult to evaluate because of their imprecision: what does "general (or special) learning capacity" mean? The problem becomes clear when reformulated in neuropsychological terms, e.g. as "Are there any special neural systems ("areas", "structures") that do the language learning and no other?" The first conjunct of this problem has a definite answer in the affirmative: the Wernicke and Broca "areas" are the ones in charge of verbal language. (However, their location varies considerably with the individual.) But the second conjunct, namely the question whether such "areas" cannot learn anything extralinguistic, is yet to be answered or perhaps even asked.

A second controversy is closely related to the foregoing and it concerns the way we learn languages (cf. Sampson, 1978; Stich, 1979; Chomsky, 1980). Is it a step by step process, essentially one of trial and error operating on a blank slate (*tabula rasa*), as empiricism maintains? Or are we born with tacit knowledge of a Universal Grammar—common to all natural languages—as the innatists (or rationalists) hold? The latter have attacked linguistic empiricism holding that it does not explain why children learn so quickly the correct grammar: if they were to learn by trial and error they would never learn to speak at all, for it would take them an eternity to discover all the rules of the language from any given set of linguistic data. True, but the falsity of empiricism does not prove rationalism—the more so since there is not the slightest evidence for the inheritability of anything beyond the mere capacity or potential to learn languages (or farming or astronomy). What we do know instead is that all normal humans share certain neurophysiological characteristics and are born into some culture or other; and that both the *neural universals* and the *cultural universals* determine the ways we learn. Some of these, which are properties or regularities of concrete systems (neural and social respectively), could be the linguistic universals that Chomsky and his followers attribute to an abstract (and yet to be discovered) Universal Grammar.

To conclude this section, I submit that the problem of language learning is a scientific not a philosophical one, and therefore neither of the two philosophies involved, namely empiricism and rationalism, is competent to solve it. We still know too little about the matter and won't learn more unless the acquisition of a first language be studied as a biosocial process, one of simultaneous neural development, socialization, and enculturation. (See Hebb *et. al.* (1971) for suggestive psychobiological hypotheses on first language learning.) We might have a satisfactory answer if the problem had been treated as one in biological and social psychology rather than in pure (nonbiological and nonsociological) psychology, let alone as a puzzle in the prescientific philosophy of mind. (See Bunge, 1983.)

3.2. *Language, Thought, and Reality*

Regardless of its unknown—though undoubtedly humble—origins, today's human language is unique. Unlike any other systems of signals, or communication methods, human language has become a *general purpose device* (James, 1890, II, p. 356). Indeed it can convey (if suitably enriched when needed) any message we may think of; it serves as a social link (to exchange information and modify other people's behavior) and it is a powerful thinking tool (as well as, occasionally, a cloak to disguise the absence of thought).

Although neurolinguists have found that language and thought are "subserved" (produced) by two distinct neural systems, there is no doubt that these systems are intimately related. (Section 3.1.) There is no speech understanding without ideation, and ideation is in turn greatly facilitated by language. The connections between the two neural systems are neural, hence material, not immaterial connections between two compartments of an immaterial mind. Some such connections are newly acquired (as when we learn to name an object), others are lost (temporarily or forever), and still others change (as when we decide to give a new idea an old name, or an old idea a new name). The set of all such connections can be summarized into a code that pairs thoughts (and images, feelings, etc.) with sound sequences or visual images.

Thought can be encoded linguistically to the point that much of our speech and writing is made up of ready made phrases expressing thoughts we have had previously. This well known fact has suggested several hypotheses concerning the role of language in thinking and behaving. One is that language guides thought and behavior, so that the latter should be explained in terms of the former (Luria, 1961). However, developmental psychologists and psycholinguists (Piaget, 1955; Slobin, 1973; Clark, 1977) have found that the acquisition of language follows that of knowledge: in the child perception, imagination and ideation precede verbal expression. (For experimental evidence against Luria's hypothesis that language regulates behavior see Bloor (1977).) What is true is that (*a*) once a language has been acquired it can greatly facilitate the learning of new ideas expressed or expressible in the language; (*b*) whereas sometimes thought appears to come in the wake of language, usually it is facilitated by the latter, which on occasion replaces it; (*c*) no language can be learned except with the help of perception, feeling, imagination, and ideation.

A second conjecture is that the child is born with "the language of thought", or mentalese, which he then learns to translate into English, Sanskrit, Maya, or what have you. The way he learns to perform such translations is by making hypotheses, testing, and revising them. These hypotheses would be of the form "$L(x)$ is true (in language L) if, and only if, $M(x)$" where M is a predicate in the mentalese repertoire of the child (Fodor, 1975). Clearly, this extreme rationalist view renders the learning of new concepts impossible; also, it is at variance with neuroscience, which shows the cortex of the infant to be poorly organized. (See Smith Churchland (1978) for further criticisms.)

A third view concerning the importance of language is the Whorf–Sapir thesis. In its extreme form it states that language is a world view imposing upon the speaker a particular way of looking at the world and therefore of perceiving and understanding it (Whorf, 1956). This is a form of idealism close to Kant's: it may be called *linguistic idealism* (Ferrater Mora, 1967), and it happens to be false. The language of every society is part of its culture and is determined by its overall level of advancement, in particular its fund of knowledge. A language is not a point of view, let alone a world view: it is subject-matter and ideologically neutral—so much so that one and the same language can serve to express mutually incompatible world views.

If a language lacks a word for the concept of atom, or of theory, this fact may indicate that it belongs to a culture whose members do not know about atoms or theories. But surely some of these people can learn to think and talk about such initially alien objects if they care to. As a matter of fact millions of Indians, Japanese and Chinese have assimilated much of Western culture and many of them have made important contributions to it—to the point that it is no longer correct to speak of 'Western' culture. What used to be given this name has become universal advanced culture, and its achievements can be communicated in many natural languages enriched with mathematics. Whatever expressive shortcomings a natural language may have are culturally conditioned and can therefore be overcome for, by and large, life molds language rather than the other way round. In particular, social status and geographical region determine the dialect the speaker will learn, not conversely. (See Trudgill, 1974.) In sum, the Whorf–Sapir thesis is false. (For further criticisms see Rosch (1977).)

A fourth well known thesis concerning the relation between language and reality is that language pictures the world: every sentence in a language would mirror a fact (Wittgenstein, 1922). A first objection to this thesis is that theories enriched with data, not languages, are supposed to represent the world. (Thus Maxwell's theory, not English, represents electromagnetic fields.) A second objection to the picture view of language is that it ignores thought. Actually we have to do with three different sets of object: facts in the external world, thoughts (brain processes), and linguistic processes (e.g. utterances): recall Vol. 1, Ch. 1, Section 3.2. And the correspondences between these sets are far from simple. In some cases a given external fact elicits no thoughts at all, at other times a single thought, and sometimes a numerous family of thoughts. Whereas some of these are verbalized, others are not; and, when they are, there is rarely a unique way in which they can be expressed. Finally there are thoughts without correlates in the external world, as well as bits of language that correspond to no thoughts at all: see Figure 3.5.

All four above theses concerning the relation between language and knowledge are false and they can be traced back to the theological assumption that wisdom is locked in Scriptures, or writings of alleged divine inspiration. This being so, wisdom can be ferreted out of words

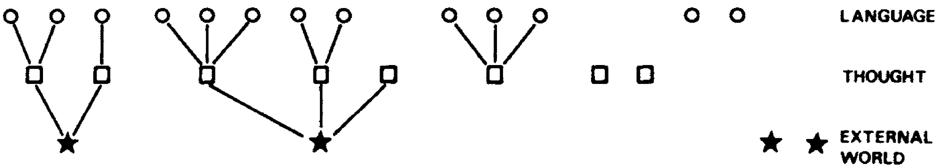


Fig. 3.5. Language, thought, and the external world. Some external world facts are not thought about, hence not described in any language, and some linguistic signs do not stand for any thoughts.

alone, be it by interpreting them correctly (hermeneutics) or by analyzing them (linguistic philosophy). Modern science rejected this approach and turned instead to experiment, theory and action. The later Wittgenstein and his disciples, as well as the German hermeneutic movement, have resurrected the biblical myth that in the beginning was the word—a counter-revolution in philosophy.

Language is both a brain tool and a social tool: it is a means for acquiring or creating knowledge as well as for transmitting it. To be sure not all knowledge is suitably communicated by word: thus motor skills and consequently much artisanal know-how is best learned by imitation. On the other hand all conceptual knowledge is communicable and thus socializable. Hence every major change in the mode of communication is bound to exert a deep influence upon the creation and diffusion of knowledge. In particular writing altered radically the mode of storing and sharing knowledge, as well as the mode of thinking: it made it possible to examine discourse in a far easier and detached way, namely by relying on visual perception. "No longer did the problem of memory storage dominate man's intellectual life; the human mind was freed to study static 'text' (rather than be limited by participation in the dynamic 'utterance'), a process that enabled man to stand back from his creation and examine it in a more abstract, generalized, and 'rational' way" (Goody, 1977, p. 37).

According to Comte every society is bound to go through three stages of development: religion, metaphysics, and science. Others have spoken of primitive and archaic knowledge, philosophy, and science. Finally others have distinguished the following stages: preliterate, literate but prescientific, and scientific. We may adopt the latter periodization, for literacy is a watershed: before it, knowledge could not accumulate beyond the storage capacity of the individual, and it could not be examined in a detached manner as if it were a material object. Besides, only a literate community can build public knowledge stores as well as collective cognitive enterprises characterized not only by their bulk but also by their variety and pace. And only literate peoples can take advantage of the division of labor between those who produce, utilize or diffuse knowledge, on the one hand, and those who handle its externalizations—librarians, book printers and sellers, telecommunications workers, computer programmers, etc.

Literacy was a momentous achievement and yet, like all progress, it had some ambivalent and even some negative effects. Its ambivalent effects were these: (*a*) it gave rise to a new caste, that of scribes, which was (and is) usually in the employment and service of the powers that be—though occasionally some members of it turn against those powers; (*b*) it gave rulers an additional instrument of administration and oppression—but one which could also be turned against them. Literacy had also some negative effects: (*a*) it facilitated the defense of the establishment and its orthodoxy by introducing the categories of holy writ (or body of untouchable belief) and or legal code (or body of rules of conduct) that survived individuals and seemed to have lives of their own above the short-lived and miserable existence of mere humans; (*b*) it reinforced psychophysical dualism, i.e. the idea that thoughts can be detached from thinkers and subsist happily ever after in a world of their own (Plato's, Hegel's, or Popper's realm of ideas). In this respect literacy and, in general, language, is like any other instrument: it can be used for good or for evil.

4. CONCLUDING REMARKS

Learning can be direct (personal) or, by means of communication, indirect (from others). Direct learning can be from experience and, in the higher animals, also from cogitation. Indirect learning can be by imitation or by instruction—or so-called 'self-instruction' with the help of books, journals, or computers. However, in social animals even direct learning, though a brain process, is done in society and largely through social intercourse.

Not only is all human learning social, but all social relations among humans are accompanied by communication of some sort or other. Therefore the study of cognition and language must not disregard their social matrix, and the study of social structure must not underrate communication. This does not entail that we are to blow up the sociality of learning or the informational aspect of sociality. We should not forget that only individuals (living in society) can do the learning, and that all means of communication are primarily means for transmitting knowledge or influencing the behavior of individuals. This caution is in order in view of the contemporary tendency to reduce knowledge, and even social relations, to information flows. There is no information flow unless there is either knowledge to be obtained or transmitted, or action to be elicited or prevented.

Mankind has evolved the most expressive and precise means of communication known to us, namely language. This is not only our most sophisticated mode of social behavior and our paramount means of communication but also a powerful thinking tool, though no substitute for thought and not more valuable than images. Though indispensable to circulate knowledge and extremely useful as a thinking tool, language must not be mistaken for knowledge, just as money must not be confused with production and trade.

This closes our preliminary discussion of the acquisition, use and diffusion of knowledge. Henceforth we shall study particular modes of cognition, in the first place perception.

PART II

PERCEIVING AND THINKING

CHAPTER 4

PERCEIVING

Perception is the basic mode of cognition. Most of the animals capable of knowing get by with only perceptual knowledge. The higher vertebrates also attain conceptual knowledge, which enriches, corrects, and guides perception as well as action.

The best known view on perception is the so-called causal theory of perception, adopted by most empiricist philosophers (in particular Locke (1689)) and all behaviorists (in particular Pavlov (1955)). According to it all our perceptions are caused by external events and more or less mirror the latter. This view is a hypothesis rather than a theory (or hypothetico-deductive system), and it is only partially true. First, it ignores hallucination and illusion. Second, it ignores inhibition, which now obliterates, now enhances the effect of sensory stimulation. Third, it ignores that our expectations, values and hypotheses condition our perceptions. (We indulge not only in wishful thinking but also in wishful seeing.) We need then a truer account of perception, one making room for the creativity of the brain and, in particular, for the intertwining of perception and conception.

The causal theory of perception has been a constant concomitant of epistemological realism. This does not entail that, by correcting it, we have to give up realism. Even though our nervous systems are not passive recorders of environmental events, they do strive to perceive them correctly, and occasionally they succeed in doing so with the help of conception and action and, in recent times, also with the help of artificial sensors and recorders. We do not construct reality every time we look at or think of it. Le Roy was mistaken to assert that *"Le savant crée le fait"*, and so are the anthropologists and sociologists of knowledge who unwittingly follow him believing that they are being faithful to Kuhn's (1962) somewhat ambiguous statements. The student of facts is given not only her brain and the cultural tradition into which she is born but also the very object of her study, namely the world—the only foundation of knowledge. What she is not given is a full understanding and control of her environment. But in a democratic society she is expected to learn something about her surroundings and herself, to make some contribution to the understanding of the world, and to have a share in the control of it.

Perception occurs both at the beginning and at the end of every cognitive process concerning reality: it supplies facts to be accounted for as well as facts testing our accounts of the former. In this sense perception is the alpha and the omega of cognition. But by the same token it does not exhaust the entire alphabet: the remaining letters symbolize so many conceptual operations. One such conceptual operation is the explicit formulation of the problem posed by the initial perceptual data—e.g. What is that?, or Why is that so? Another is the formation of tentative hypotheses designed to solve such a problem. The test of such hypotheses joins conception with perception. Cognitive success lies then not in reducing conception to perception (radical empiricism) or in discarding the latter (radical rationalism) but in the alternation and confrontation of perception and conception.

1. FROM SENSING TO PERCEIVING

1.1. *Sensing*

All organisms detect some environmental changes, but only the animals endowed with special detectors, namely sense organs, may be said to sense them. Thus whereas the octopus can see its immediate environment and discriminate among its components, presumably the sea urchin, which lacks sense organs, only detects changes in stimuli (von Uexküll, 1921).

Sensation is the specific function or activity of neurosensors such as the retina and the taste buds. Neurosensors are ordinarily multicellular systems specialized in detecting stimuli of certain kinds coming from the environment or from other parts of the animal. Like microscopes and photovoltaic cells, neurosensors filter out some (actually most) of the stimuli impinging upon them, and they accentuate the spatial heterogeneities and temporal changes in the environment, i.e. they enhance contrasts. For this reason they are similar to tuned filters (Ratliff, 1965, 1971).

Both the filtering out of stimuli and the enhancement of contrasts are results of inhibition. Thus when looking at a thing our success in seeing it depends not only upon the illumination of the external object but also upon the inhibition of the neurons in the visual cortex surrounding those that have been activated. (Thus, as Leonardo noted 500 years ago, when looking at a printed page you see all the printed symbols at a glance but, in order to make out the words, you must focus on a small region of the page. In modern terms, in order to perceive what you detect you must inhibit your detectors.) The same holds for the other senses, in particular hearing and touch. This feature, lateral inhibition, is a peculiar property of the nervous tissue totally ignored by S-R psychology. It had been discovered by Mach one century ago, later forgotten, and rediscovered by Békesy (1967). From here on the minimal possible psychology is S-I-R, not S-R.

There is agreement that our sense organs are adaptations, i.e. results of genic variations followed by natural selection. However, such adaptation has not consisted in a gradual improvement of all the senses. Thus it is likely that our remote ancestors were capable of hearing, tasting and smelling better than we do, but we seem to be able to see better than they did. In addition to evolution we must take development and training into consideration. Thus a child's sensory discrimination is poorer than a normal adult's, and that of a technician or an experimental scientist (in particular an optician) is far superior to that of a layman.

Even so, our best senses have well known limitations. For example, if a thing moves very fast we see it as blurred or even not at all; and if a multicolored disk rotates rapidly we see only a homogeneous grey patch. And, no matter what the sensory attribute may be, we cannot distinguish without confusion more than about six different stimuli—e.g. visual position, or pitch, or salinity. In other words, our channel capacity is rather poor and is fairly uniform over the entire sensory domain (Miller, 1967). However, sensory discrimination can be educated to attain a remarkable sensitivity in musicians, microscopists, color experts, wine tasters, and other sensory specialists.

The phenomena of filtration, contrast enhancement, and confusion (poor discrimination) show that the vertebrate nervous system is not so much an information processor as a sink. When impinging on the central nervous system some stimulation gets lost in the following ways: (*a*) all sense organs are highly selective filters (e.g. we do not see infrared radiation or hear ultrasound); (*b*) the excitation reaching a point in the nervous system is not propagated but circumscribed (lateral inhibition); (*c*) excitations are usually dampened; (*d*) the various internal and external stimuli often cancel each other (compensation). All this shows that the nervous system is not at the mercy of its external or even internal environment. It is permanently active and it can do its job because of the above properties. In fact filtering cuts out much disturbing excitation, such as the infrared waves generated by our own body. Inhibition improves the signal/noise ratio (e.g. yielding enhanced contrast). Damping protects us from incessant environmental shelling. And compensation discards much useless (together with some useful) information. By all these means the nervous system avoids information overload.

Accordingly the conception of the nervous system as an input-output box describable by a simple systems theory, such as information theory, is basically mistaken. Stimuli do not always excite, and those which do, do not always produce responses. On the contrary, all stimuli produce some inhibitions, and in the higher vertebrates much neural activity has no behavioral issue. (The typical mental activity is of this kind.) Therefore the S-R model is basically false. Worse, it has inhibited the investigation of inhibition and of spontaneous neural activity.

A similar criticism is applicable to the so-called ecological theory of perception (Gibson, 1966, 1979). According to this view, information resides in the environment: the organism picks it up ready made, processes it, and uses it, in analogy with the way it processes food. A first objection to this view is that information does not lie around us: we have to produce it, either by transforming sensation into perception, or by generating it internally (in ideation). Even the information-carrying signals impinging upon our neurosensors, such as the images on a TV screen, have to be "read" by the nervous system before they become information. Secondly, ambiguous figures, such as the Necker cube or some of Escher's *trompe l'oeils*, elicit mutually incompatible perceptions: they do not yield unambiguous information. The subject constructs alternatively mutually incompatible information packages out of one and the same stimulus. Thirdly, the subject, far from being a virgin film, has certain schemata (Bartlett, Piaget, Neisser) that allow her to perceive certain things rather than others, as well as to explore her environment searching for certain stimuli (e.g. moving objects) in preference to others.

In sum, the animal detects (some) environmental changes impinging upon it (i.e. it senses stimuli) and perceives ("interprets") some of them: it never picks up ready made information lying about it. Information about a stimulus equals perception of the stimulus. An unperceived stimulus carries no information; an incorrectly perceived stimulus carries wrong information; and an ambiguous stimulus elicits the generation of mutually incompatible pieces of information. In all such cases *information is generated by the organism* not just received and processed by it. Hence we should speak of *stimulus pickup and transduction* instead of *information pickup*, and we should drop the simplistic assumption that information theory is the answer to the problem of perception. But we are encroaching on the subject of the next subsection.

To conclude this subsection. The neurosensors supply the raw material of our knowledge of the external and internal worlds. Without them we would know nothing, with them alone we would know nothing either. This is not philosophical dogma but a result of empirical research. In fact (*a*) subjects cut off from the world, in experiments on sensory deprivation, become spatially and temporally confused, and they start to hallucinate; and (*b*) subjects with their neurosensors intact but damages in the sensory areas of their brain cortex are incapable of processing the incoming stimuli: they can detect but not perceive (agnosia).

1.2. *Perceiving*

Physics and chemistry explain what happens in the eye from the moment light enters into it till it strikes the rods or cones in the retina. (Hearing is parallel.) From then on neurophysiology and psychobiology must take over. Thus the containment of the excitation, i.e. the inhibition of the cells surrounding those that are excited, is a typically biological process. The continuity of the visual world despite our blinking every few seconds may be an effect of a neural inhibition generated by the brain, that diminishes the effect of the blinks (Volkmann *et al.*, 1980). The discarding of small differences (as in pattern recognition) is the job of specialized (feature) cells in the primary visual cortex. And our perceiving distant objects almost as large as close ones, despite the fact that they project a smaller image on to our retina, is presumably a result of the activity of large neuronal systems in the secondary or tertiary visual cortex. All such phenomena illustrate the difference between sensation (or recording) and perception (or elaboration of sensation).

Whereas sensation is localized in the neurosensors, perception is distributed throughout large regions of the brain. Recently developed techniques for imaging and measuring local cerebral blood flow, metabolism, and other processes, allow neuropsychologists to locate the areas of the brain that are activated by a specific sensory stimulus. (Cf. Phelps *et al.*, 1981; Greenberg *et al.*, 1981.) One result is that the images of the brains of subjects given somatosensory, visual, or auditory stimulation are quite different. Another is that they have a significant overlap, suggesting that some neuronal systems are activated by sensory stimuli of any type. This may explain why our perceptions are strongly colored by our memories and expectations and why they are accompanied by some movement or other.

We have no direct knowledge of the external world, for what is sensed is some event at the tip of a sensory nerve. The attribution of this even to an environmental item is a brain construction—either a perception or a hypothesis—and one that is sometimes correct and at other times incorrect. Thus a dog wounded by a stick does not snarl at its wound but at its attacker. We share with many animals "the power of imputing otherness and beyondness" (Lovejoy, 1929). The constructive nature of perception is further shown by the fact that we do not have to see a familiar thing in its entirety to recognize it: seeing only part of it, at a given angle, often suffices. (If it did not we would never recognize anything, for we cannot see any thing in its entirety.) In other words, neurosensors sample (and analyze) clusters of features, which are completed and integrated by the brain during perception.

Perception integrates sensation, is accompanied by behavior (e.g. eye movement during scanning), and is guided or misguided by conception. Indeed it has been shown experimentally that, if motor activity is impeded, e.g. by immobilizing the subject's eyes or limbs, then visual perception stops altogether. And if a subject wears a prism that reverses right and left, or top and down, and moves around, eventually the world "rights itself" for him: his plastic brain establishes new connections. In short, the way we perceive depends on the way we behave and conversely. (Cf. Taylor, 1962.)

Likewise what we perceive depends not only upon the stimulus object but also upon our knowledge and expectations. In short, in all likelihood there is no pure perception untainted by ideation and behavior. This result of recent physiological psychology had been anticipated by a number of philosophers. In particular Bacon, Kant and Whewell held that the way we perceive an external object is partly determined by our knowledge of it. Thus when I report that I am seeing Venus, I am drawing on my fund of knowledge: another person presented with the same celestial object might claim that he sees a star, a UFO, or an angel. This being so, it is very likely that our ancestors, even the recent ones, did not perceive the world exactly as we do. However, it was roughly the same world as ours.

Consider listening to speech or reading. Normally "concrete" words, such as 'run', trigger images, whereas "abstract" words, such as 'commutativity', trigger concepts or symbolic diagrams representing concepts. And when hearing or reading a new word either we form neither image nor concept, or we guess the correct image or concept from the function the word discharges in the sentence: i.e. we conjecture its meaning from certain contextual clues. In short, speech perception and reading are highly constructive processes involving sensing, perceiving, imagining, remembering, conjecturing, and expecting. (See Carterette and Friedman (Eds.), 1976.) The reading of musical scores, diagrams, graphs, blueprints, and the like, must be parallel.

Percepts do not appear automatically and immediately upon sensory stimulation, but are constructed or synthesized (Selfridge and Neisser, 1960). This constructive view of perception has been confirmed by neurophysiology. In fact, we know from the work of Hubel and Wiesel (1962), Evans and Whitfield (1964), and others, that certain cortical cells specialize in recording single features (e.g. edge, inclination, or color), and that their output is synthesized by further neural systems. We also know that there are no isolated neural systems, so it is conceivable that the activity of the perceptual synthesizers is controlled by memory, expectation, and ideation. See Figure 4.1.

Throughout the history of philosophy and psychology two major views on perception have clashed: those on perception as recording (or passive models) and as constructing (or active models). Naive realism and empiricism have held the former, whereas idealism has held the latter. The passive models are the slate and the filter: according to them all the nervous

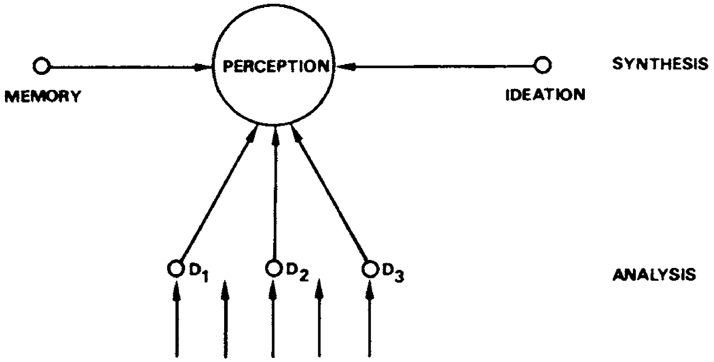


Fig. 4.1. Cells D_1, D_2 and D_3, of different kinds, respond selectively to different features of one and the same stimulus, and their outputs are synthesized by a neuronal system (the perceptor). Some of the features of the stimulus are either not detected at all or suppressed. The synthesis is not a neutral summation but is influenced by the current state of the whole brain, in particular its memories and expectations. Thus when we say "I know this girl" when meeting a particular girl, we mean that we have recognized her: our perception of her is "colored" by our memories of her and perhaps even our expectations about her.

system does is to record and select the incoming stimuli. A deterministic automaton with or without output would be a good mathematical representation of such models.

Obviously, no passive model harmonizes with neuroscience, in particular the findings about lateral inhibition and the spontaneous ongoing activity of neurons and neuronal systems, in particular the integration of the outputs of specialized detectors (recall Figure 4.1). These findings support on the other hand some active or constructive model, according to which perception is a brain activity resulting from the joint action of input signals and internal processes. As a matter of fact it has been known since Helmholtz (1879) that the senses, far from giving us a *picture* of the world, give us only *signs* of it, which signs must be "interpreted" by the brain to become cognitive items. In short, we do not just copy or reflect but construct. However, this does not endorse idealism: we do not construct the world but only map it.

Every animal is equipped not only with a number of sensory systems but also with certain predispositions or propensities to use such systems in certain ways rather than others. Thus we are predisposed to paying attention to moving objects and shrill sounds; to listen to and imitate human speech in certain languages rather than others (similarly most song birds imitate the song of their conspecifics even when they are exposed to the song of birds belonging to other species, or to music); we notice items of certain kinds more readily than others (e.g. expecting parents tend to see pregnant women and babies everywhere), and so on. These perceptual propensities are inherited or acquired properties of our brain and they develop with it: they allow us to have certain experiences and are in turn modifiable by the latter. Psychologists call such predispositions perceptual *schemata* (Bartlett, 1932; Piaget, 1952; Neisser, 1976).

A perceptual schema is neither a thing (e.g. a neuronal system) nor a process in a thing. It is a cluster of properties of a neuronal system and, more particularly, a bundle of propensities for it to act in a certain manner in the presence of external or internal stimuli of a certain kind. Presumably such propensities can be quantitated as probabilities, so that all models of perception should be stochastic models. However, this is just conjecture: we still do not have adequate models allowing us to explain how a person in a certain state sees a chair as a piece of furniture, and when in a different state as a weapon or as a piece of firewood. The most sophisticated models of perception, those of Hoffman (1966), are not stochastic but use Lie groups, and they do not involve any neuroscience. A characteristic hypothesis of such models is that a global pattern is the result of an iteration of Lie operators—rather than, say, the activity of a neuronal system activated by a number of sensory analyzers or detectors.

Adequate models of perception can be produced only by neuropsychology (or physiological psychology): they will not come from pure psychology, let alone from philosophical psychology, which has been at it for over two millennia without ever getting hold of that which does the perceiving, namely the central nervous system. A good example of the failure of armchair speculation on perception is the "new theory of perception" advanced by Brown (1977). This view is claimed to elucidate Hanson's idea that scientific percepts are theory-laden. (Actually this idea goes back to Whewell, who took it from Kant, who knew about Bacon's *idola.*) The gist of this view is that the distinction between *seeing* (e.g. a shadow) and *seeing that* (e.g. that the shadow has been projected by a person) is that in the second case the observer takes in "the meaning of the observed situation" (Brown, 1977, p. 87). The author calls this 'significant perception' in contrast with sheer perception (actually sensation). But of course there is no *theory* proper (i.e. hypothetical-deductive system) here. There is not even analysis, for Brown does not elucidate the key notion of "meaning of a situation". And there is no novelty, for the whole thing boils down to the old finding that we perceive what we are prepared to perceive. The challenging task is not to use ordinary language expressions to describe once more the difference between untutored sensation and educated perception, but to build a psychobiological theory couched in mathematical language.

Understandably, most of the research in perception has been concerned with vision. (If we dwelt in deep caves we would presumably have specialized in tactile and auditory perception.) However, not all perception is of some external object: there is also perception elicited by bodily processes. Thus we perceive the pressure of one hand rubbing the other, and we perceive (painfully) the thorn in the toe or the ulcer in the duodenum. Besides, the brain can generate perceptions that have no stimulus object, whether external or internal: such are hallucinations and imaginations. (The only difference between these two is that the former falsely imputes some external stimulus object.) We can imagine pictures, sounds, bodily processes, and feelings. We can do so at any time, while awake or in sleep, working or daydreaming, in normal circumstances or under the action of drugs or of hypnosis. (There is no such thing as a hypnotic trance: the good hypnotic subject is highly suggestible and imagines whatever is suggested by the hypnotizer: Barber and Wilson, 1979.)

Imagination is a mode of cognition without which nothing great is ever achieved in any field. True, the products of imagination must be checked for truth or efficiency unless we are bent on creating a work of art. But the same holds for ordinary perception and conception as well: we are not to trust any single cognitive operation. Moreover such checking is never automatic and infallible: it takes a lot of imagination to produce firm evidence for or against the works of imagination. Only the empiricist obsession with empirical data could make us forget for a moment that the most valuable data are those produced with the help of imaginative means such as sophisticated experiments inspired by sophisticated theories. Scientists and technologists do not shun imagination: they cultivate it and keep it under control.

Imagination takes part also in the intuitive or ordinary knowledge of other minds known as empathy. Empathic understanding, which we share with many animals, is nothing but imagining what goes on in someone else's brain, or "putting oneself in the other's shoes". It is compassion, coperception and coideation. Empathy uses (ambiguous) clues and (fallible) conjectures relating overt behavior to mental states, e.g. the shedding of tears to sadness, and the shuffling of feet to embarrassment. Like every mode of perception, empathy is fallible: the illusion of infallibility comes from our noting its hits and disregarding its misses. "Lovers, who are notoriously supposed to be empathic, as frequently seem to misunderstand as understand each other's motives and acts; and what they normally do, perhaps, is wilfully misunderstand their existing differences, so that they create an illusion of unanimity of opinion and feeling" (Ellis, 1963, p. 119).

What about space and time: are they perceptible? Although many psychologists write about the perceptions of space and time, these are imperceptible: only (some) events are perceptible (under special circumstances). What we can perceive is not space but events that hold spatial relations; nor do we perceive time but events in temporal relations. Thus we hear one noise to the right of another, and perceive the former before the latter, but "to the right" is just as imperceptible as "before". In general, we perceive neither things nor properties nor relations: we perceive only some events, which are changes in properties of things or relations among things. If there were changeless things we would not notice them.

Finally, can we perceive external objects without the help of the senses:

i.e. is there extrasensory perception (ESP)? In the context of traditional psychology, which ignores the nervous system, this question must be regarded as open, for one might hope (or fear) to get eventually some firm evidence (other than statistical mistake or deception) for the hypothesis of ESP. Likewise before the nature of meteorological events was understood one could hope or fear for a storm to strike out of the blue. We now know that the perception of a stimulus object (unlike imagination and hallucination) is the terminus of a process that starts at some nerve ending or other, and ends up in a secondary or tertiary sensory area in the cortex. This being so, clairvoyance and other alleged ESP phenomena are as illusory as running without limbs or digesting without guts.

It is a firm finding of physiological psychology, and a principle of our epistemology, that all external perception is mediated by some sensory system. Put negatively: there is no ESP and there can be none. In particular there is no direct communication among animals at a distance: all communication proceeds *via* physical signals (light, sound, touch, etc.). And what mind reading there is consists in perceiving and interpreting the movements or signals produced by another animal. Thus you are reading my mind as you read these lines.

2. OBSERVING

2.1. *Direct Observation*

There is an important difference between spontaneous and directed perception, or sheer perception and observation. In the former case we look *at* things, in the latter we look *for* them. Directed perception, or observation, is purposive, active, and deliberately selective. A dog and a human being watching the passing scene are onlookers; a dog looking for a bone, and a human being looking for a friend, are observers. The difference is one of psychological "set" (*Einstellung*): different neuronal systems are presumably involved.

Because observation is purposive and selective perception, it can be educated. (Recall Leonardo's injunction to his students: *Saper vedere.*) In fact the training of a young hawk, a tiger cub, or a human infant, are largely education in observation skills: in focusing on certain things or traits while disregarding the rest, or enhancing the effect of certain stimuli while inhibiting others. (Békesy (1968) stated that "a good observer is an observer who learns to inhibit". But so is an observer incapable of perceiving due to bias or to unwillingness to face reality. Thus Tolstoy's Karenin "was unwilling to see anything and so saw nothing".) There is little doubt that our children, who are exposed from birth to a ceaseless stream of pictures and noises, are better lookers but worse listeners than our fathers. In the case of humans this education is not only sensory-motor but also intellectual. We perceive almost exclusively that which we have been trained to perceive (including self-training). If our brain has been soaked since childhood with witches, devils and conspirators, we are bound to look for them and likely to find them. If on the other hand we have had the benefit of a scientific or technological education we shall tend to look only for concrete things that can be accounted for in naturalistic terms.

Two persons, a farmer and a scientist, see a brilliant thing now hovering above them, now moving swiftly away, and neither can identify it immediately. If the farmer has heard about angels or UFOs (first "seen" in 1947), he may believe he is seeing one of either. The scientist, on the other hand, is likely to identify the object as an aircraft, a plasma, or Venus. Philosophical subjectivists notwithstanding, both subjects deal with the same object perceived differently. This holds not only for brilliant flying objects but for almost everything: people with different backgrounds, cognitive attitudes, and value systems, are bound to see the (only) world there is in more or less different ways.

Some philosophers, in particular Whewell (1847), Duhem (1914), and Hanson (1958), have stated that experiences like the above show that observation, far from being neutral or uncontaminated by theory, is penetrated by it. (Hanson introduced the expression *theory-laden*, which is incorrect, for most observers do not know any theories proper. The expression *hypothesis-driven* seems preferable.) There is a grain of truth in this: we do tend to observe what we expect to perceive, we often do miss what we do not anticipate to perceive, and all our expectations are determined by our fund of knowledge (and superstition) as well as by our value system (or interests). However, that grain of truth should not be puffed up to the point of epistemological relativism and subjectivism. The observations on the brilliant flying object reported by the farmer and the scientist are not equally correct: they are mutually incompatible, both can be checked for objective truth, and it so happens that the checks favor the scientific version.

True, observation is hypothesis-driven, particularly in the sciences, because it is selective. The myth of immaculate observation is all but dead, even in predominantly observational disciplines such as ethology (Hinde, 1970; Beer, 1973). But, if scientific, rival hypotheses are checkable. And, once checked (against data or theories or both), they are bound to be shown unequal in worth (truth value, generality, or depth). The same holds for data: many of them can be replicated, and all of them are (in principle) checkable by different means or methods. Consequently even if all of them are hypothesis-driven, many of them are *invariant with respect to changes in the driving hypothesis*. So, such data are in effect uncontaminated by any particular hypotheses. This is the case with many data stemming from direct observation, i.e. observation unaided by any instruments, such as those gathered by a census bureau. On the other hand indirect observation, i.e. observation conducted with the help of scientific instruments, is not only driven by hypotheses but is also hypothesis-dependent, i.e. its worth depends in part upon the hypotheses involved in the design and operation of such instruments. We shall return to this in the next subsection.

What is the epistemic value of observation? According to empiricism all and only observation supplies genuine knowledge: i.e. observation is necessary and sufficient for it. Epistemological analysis shows that observation, though necessary, is not sufficient, for all observation is hypothesis-driven, and all indirect observation is hypothesis-dependent. Thus finding a fossil bone is not, in itself, a scientific discovery: many people had found, and discarded or misinterpreted, fossil bones before evolutionary biology made it possible to "interpret" them as remains of organisms that lived long ago. An empirical scientific discovery is a finding together with a theoretically plausible account of it. Hence two rival accounts of one and the same finding must rate as different discoveries—at most one of which can be correct. Thus Columbus and Vespucci, though interpreting roughly the same geographical findings, made different discoveries: in fact only Amerigo Vespucci discovered America. Likewise although the ancient Greeks knew about amber and loadstone, they did not discover electricity or magnetism; and the discoverers of DNA did not realize that it is the genetic material, and so did not discover that DNA is the carrier of heredity.

Observation, though insufficient, is necessary for genuine knowledge of fact. True, we cannot observe directly neutrinos or synaptic boutons, any more than live dinosaurs or prehistoric events. But we do make certain observations of related things that motivate and check the hypothesis of the current or past existence of unobservables. On the other hand there is no observational evidence for the hypothesis that there are disembodied entities, such as ghosts and ideas in themselves: hence this hypothesis does not count as genuine knowledge. The need for observation in gaining factual knowledge is a tacit postulate of science and technology. (Even the highest level scientific theories, such as quantum mechanics, contain constants or parameters whose values can be known only through measurement. And all measurement involves, at some point or other, direct observation.) Therefore we incorporate that principle into our epistemology: *All knowledge of factual matters consists of, or involves, some direct observation.*

The historical sciences, such as cosmogony and geology, archaeology and human history, are not exempt from this condition. True, historians cannot observe events in the remote past: most of their observations are limited to remains, such as tools and written documents. They must conjecture the possible use of tools and, if they proceed scientifically, they will make replicas and try them out; and they must be able to interpret the documents and try and check them with other findings. In all cases, regardless of the amount of speculation historians may embark on, they have to make some contact with observation if they wish to substantiate their hypotheses. The same holds for all the other historical sciences, from cosmogony to evolutionary biology: there is always some observation involved in them, whether at the beginning or in the middle or at the end of a research cycle. A field of inquiry that involves no observation at all is either strictly formal (i.e. logical, mathematical or semantical), or purely speculative—and therefore nonscientific.

2.2. *Indirect Observation*

Not all facts are directly observable. By far the greatest number are not: think of electrons or quasars, of genes or nations, of atomic collisions or historical events, of extinct organisms or neural processes. All such facts go beyond appearances and therefore constitute a glaring refutation of phenomenalist epistemologies and metaphysics.

Suspecting that reality encompasses more than appearance, scientists have devised literally thousands of means for extending the reach of our senses. Strictly speaking only a few such means—e.g. the scale, the microscope and the telescope—extend the scope of the human senses. Others, such as the magnetometer and the pH meter, the voltmeter and the interferometer, the Geiger counter and the EEG, are not extrasensitive eyes, ears, tongues, or skins, but tools without biological precursors. However, all of them supply directly observable facts, such as clicks,

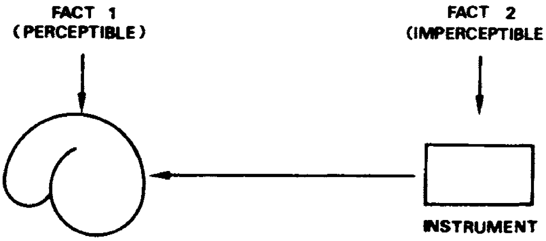


Fig. 4.2. Direct and indirect observation.

coincidences between pointers and divisions on a dial, or computer printouts. The mode of inquiry employing such means may be called *indirect observation*. See Figure 4.2.

A short and haphazard list of such means should show how much farther than our senses they go, as well as how hypothesis-dependent they are. Radiotelescopes are indispensable to study radiostars. Astronomers hope (or dread) to discover black holes either by looking for very luminous stars (binary pulsars?) that seem to revolve around nothing, or by observing the radiation emitted by objects in the process of being swallowed by black holes. Neutrino detectors allow one to "look" into the interior of a star, where they are produced, for starlight comes mostly from the thin stellar atmosphere, and neutrinos can traverse it almost undisturbed. (Incidentally, such observations have recently upset the accepted theories of the structure of stars.) Field-ion microscopes and scanning transmission electron microscopy produce direct images of individual medium to heavy atoms and molecules. (The classical techniques supply only images of average structures.) And there are several techniques for detecting what goes on in the living human brain (e.g. microelectrodes and EEG) and even for imaging such processes—e.g. computer assisted tomography (CAT scan) and positron emission tomography (PET scan).

As with direct perception, instrument assisted detection is of events or changes, and the detection itself is a change in a thing, namely the detector. To be sure the thing detected may appear to be in a stationary state, as is the case with the printed page you are looking at right now. But this is only appearance: in actual fact this page, while legible, is being bombarded by myriads of photons, some of which are absorbed, the rest reflected, and all of which affect this sheet in some manner, e.g. by warming it up and speeding up some chemical reactions in it. In short, only events are detectable.

Every means of indirect observation is conceptual as well as material, for it involves not only a material instrument, such as a scale or an amplifier, but also a set of hypotheses employed in its very design and operation. Thus, the detection of radio waves employs electromagnetic theory; X-ray and neutron crystallography involve not only pieces of apparatus but also the theory of diffraction; electrophysiology uses not only electrodes and meters of various kinds but also scraps of physical and biological theories. In many cases the theory itself suggests the very existence of the thing to be detected as well as manners of detecting it. (Thus a certain technique for detecting the elusive neutrino is based on calculations showing that, when the nucleus of gallium 71 absorbs a neutrino, it turns into the nucleus of germanium 71, a different element.)

Without such theoretical components the raw data supplied by indirect observation would be unintelligible and therefore useless, for they bear no obvious relation to the facts that are being observed. Thus a layman would not know what to make of a sequence of clicks in a Geiger counter, or of an X-ray diffraction pattern, or of a collection of spots on a sheet of chromatographic paper. Every indirect observation requires an "interpretation" of the data, which "interpretation" consists of hypotheses about the nature of the events of interest and their interaction with the detector. This intervention of hypotheses in observation is not limited to natural science, medicine and engineering, but occurs in all fields of research. Thus sociologists use social indicators, and economists economic indicators, to detect changes in society, and prehistorians examine artifacts, bones, and inscriptions to "read" the evolution of mankind. In all such cases certain hypotheses—that will be called *indicator hypotheses*—link the available data with the events of interest. (More on indicators in Ch. 11, Section 2.2.) In short, every means of indirect observation, be it of microevents such as atomic collisions, or of macroevents such as economic crises, involves a theoretical component—or, as an empiricist would say, is "contaminated" by theory.

Indirect observation has advantages and disadvantages with respect to direct observation. The advantages are obvious: (*a*) it enables us to go far beyond appearance; (*b*) because it employs impersonal instruments free from conceptual or moral bias, and yielding data that can be perceived by all competent observers, it is less subjective than direct observation. The disadvantages are equally evident: (*a*) some instrumental probes disturb the probed object (particularly in microphysics), and all instruments are likely to produce "artifacts", i.e. effects of their own (e.g. halos and stains) that the observer may mistakenly impute to the observed object; (*b*) the data resulting from indirect observation are just as fallible as their theoretical components. Not surprisingly, if one wishes to get more one must be prepared to pay more. However, the cost (i.e. error) may be lowered with improved instrumentation and theory.

As for bias caused by wishful thinking or dishonest intention, indirect observation is neither more nor less susceptible than direct observation. One can lie not only in reporting sightings with the naked eye but also with a camera—e.g. by changing the sequences of frames, or taking pictures from appropriate angles or under suitable illumination—or by splicing tapes or simply by selecting the favorable data and eliminating all the others. However, it is obvious that lying with cameras (or meters or computers) goes down better than without them. Hence the methodological and ethical precautions are more necessary in the laboratory than in the field or the street, the more so since all results of indirect observation must be doctored in some way or other.

Advanced science has little use for raw data: the latter must always be corrected for a number of effects, and sometimes (as in the case of outlying figures) thrown away altogether. The line between standard data correction and finagling, between honesty and dishonesty in reporting results of observation, is not always a clear one. Newton has been accused of smudging astronomical data and Mendel of ignoring certain results of his hybridization experiments. Morton has been shown to shift criteria, and omit or miscalculate some data on cranial capacity for harboring "an a priori conviction of racial ranking" (Gould, 1978); and Millikan discarded some values of the electric charge of the electron for failing to square with his hypothesis (since amply confirmed) that all electrons have the same charge (Holton, 1978). In all these cases preconceived ideas—sometimes true, at other times false—had distorted the presentation of data. Nor are such cases exceptional: they occur all the time in ordinary life and in the laboratory. Thus if we are used to seeing A's but not B's under certain circumstances, or if we assign A's a greater probability or utility than B's, we may miss the latter when they occur or confuse them with A's—or may be inclined to report only A's for fear of being taken for fools. (For an experimental study of biasing factors in discrimination problems, see Swets (1973).)

Not all observers are equivalent in all respects—*pace* Ziman (1979). Therefore we should not expect that different observers will always produce the same reports on one and the same group of facts. Two observers may be inequivalent because they are not in the same state of motion, or because they use different equipment, or operate the same equipment in different ways, or under different circumstances. Or they may be inequivalent because they handle their data (e.g. round off, average, or discard figures) in different ways. Or because one of them is gifted or lucky enough to notice something missed by the other, or because one of them is wedded to a certain hypothesis whereas the other is indifferent to it. (Recent examples of the effect of theoretical bias have been seen in the search for gravitational waves, black holes, and quarks.)

Even the most competent and lucky of observers is faced with the inherent limitations of every means of observation. All optical instruments have a limited resolving power due to the diffraction of the impinging light. Electronic and protonic microscopes possess a far greater resolving power but have their own limitations. Astronomers are up against the limits of their instruments as well as of the observation sites: clouds and turbulent currents, not to speak of smog, will produce a diffuse and constantly fluctuating image because of the variation in the atmospheric refraction index. Obviously, such limits to the accuracy of observation are limits to knowledge. However, some such limits can be overcome in principle and possibly in practice. For example, telescopes installed in artificial satellites or on the Moon will be free from atmospheric nuisances, and all kinds of electric meters are essentially free from thermal noise when working at extremely low temperatures.

We summarize the above discussion on the reach and limits of observation in the following *postulate. For every fact f* (thing or event) physically accessible to cognitive subjects (*a*) *there is* (actually or potentially) *a means m of observation such that f has observable traits relative to m* (i.e. such that *m* helps detect some properties of *f*), *but (b) there is no means allowing one to observe all of the traits of f with complete accuracy.* In short, the means of observation can all be improved but never be rendered perfect. (*Notes*: (*a*) the physical accessibility condition excludes most events in the remote past, as well as all future events and those outside the observer's light cone; (*b*) the limits on observation have nothing to do with Heisenberg's indeterminacy inequalities, which only limit the applicability of the classical concept of a corpuscle.)

3. MAPPING REALITY

3.1. *Perceptual Mapping*

Perceptions, whether spontaneous or sought, do not pile up haphazardly but tend to get organized into systems. Such systems of percepts are called *cognitive maps* or *internal representations*. Some of them represent our environment (*environmental maps*), others our own internal states (*body maps*). These maps are continually updated, they summarize our perceptual knowledge of our surroundings and of ourselves, and in turn they help us take care of ourselves, move around, and act: see Figure 4.3.

We do not know exactly what animal species are capable of forming perceptual maps. All we seem to know is that, in order to map or model its surroundings, an animal must possess a nervous system capable of (*a*) *analyzing* or breaking down into distinct units whatever it perceives, (*b*) *recognizing types* or kinds into which such units can be grouped, and (*c*) *organizing* what it perceives into a whole, i.e. synthesize or integrate the products of analysis and type recognition. Presumably, it takes a rather complex nervous system to accomplish all this—say, that of a fly at least—but we do not know exactly what it takes.

According to psychoneural dualism, cognitive maps are immaterial: they are "in the mind". So, by hypothesis subhuman animals would not be entitled to forming cognitive maps. On our biological perspective a cognitive map is a pattern of neural activity in perceptual and motor centers. In particular, it has been conjectured that the internal representation of an object seen by a normal human subject is a sort of ring formed by sensory memory traces and motor memory traces (Noton and Stark, 1971). As a subject sees a thing for the first time, she forms a scan

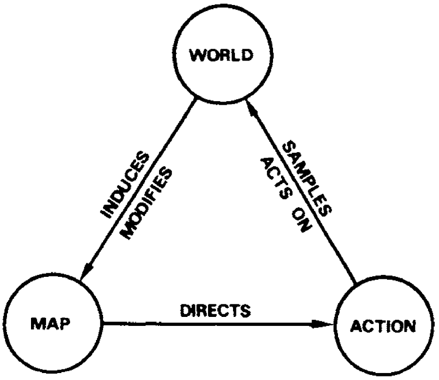


Fig. 4.3. Cyclic model of perception (adapted from Neisser, 1976, p. 112). External objects stimulate the formation of cognitive maps in the brain. These in turn guide locomotion, exploration, and other activities that modify the environment, which changes call for revisions in the cognitive maps, and so on. Caution: both maps and actions are part of the subject, and while both must be conceived as biological processes, the world is a thing.

path of that object. This scan path consists in a chain composed by a visual feature followed by the eye movement required to reach the next visual feature, and so on. When the subject encounters the same object again, she recognizes it by matching the incoming visual stimuli with the "feature ring". Thus once constructed each such ring guides the future perception of similar objects. Each subject constructs her own feature ring, which is fairly permanent. This hypothesis may not be accurate but it matches the available experimental data—particularly all the data concerning the strong linkage between perception and motor behavior—as well as the general psychobiological orientation. It is incompatible with psychoneural dualism—a definite virtue in our view—as well as with the Gestalt hypothesis that internal representations are built in one piece and instantly.

(One way of elucidating the general concept of a cognitive map is as a function from some events in the perceptual field into neural events. More precisely, consider a concrete thing m and take the set M of events happening in m at a certain level L of discrimination—e.g. the collection of all the photons of a certain frequency bouncing off m. Further, call M_L the subset of M formed by the events above a certain level L of discrimination. Then a *cognitive map* of the thing m is a function f from a subset P_L of M_L to a set N of neural events in some subject, i.e. $f: P_L \rightarrow N$, that preserves some of the relations and operations among members of P_L. A perfect cognitive map of m is of course one that (*a*) mirrors every event happening in m, i.e. for which $P_L = M_L$, and (*b*) preserves all of the relations and operations in M_L. We may safely assume that no such perfect map is possible.)

Obviously, the cognitive maps of different individuals, even of the same species, are not identical: each of us perceives herself and the world in her own way. Still, there must be commonalities due to our common ancestry and our shared environment. In general, the individuals of each animal species would seem to map their surroundings in a species-specific way: the octopus in one manner, the owl in another, and so on, as proposed by Uexküll (1921). (Actually Uexküll held that each animal *constructs* its own environment in a species-specific way. This is either sloppy language or subjective idealism. All we do is map our environment, and use such maps to navigate in it and modify it. More on this subject in Ch. 15, Section 2.2.)

In the recent psychological literature the expression 'cognitive map' is usually understood in the restricted sense of a *spatial* map or schema allowing an animal to orient itself in ordinary physical space. Thus O'Keefe and Nadel (1978) have conjectured that all mammals are born with a framework, located in the hippocampus, where they can store all the information concerning places and spatial relations that they gather in the course of their lives. They further conjecture that, "When there is a mismatch between some sensory input to a place representation and the predicted sensory input, a set of *misplace* detectors is triggered, the output of which activates and directs the motor systems involved in exploration. The behavior is directed towards the incongruence and new information can be incorporated into the map as a result of it" (p. 94).

That some animals form spatial maps of their environment, and that such maps guide much of their behavior (approach, avoidance, exploration, etc.), seems beyond reasonable doubt. What may be disputed is the thesis of Descartes and Kant, revived by O'Keefe and Nadel (1978), that all mammals are born with a *Euclidean* framework. In fact it has been known for some decades that (*a*) none of the various perceptual (visual, auditory, tactile) geometry is Euclidean—e.g. some parallel straight lines are seen as curves, whereas some curves are perceived as straight parallel lines; (*b*) children learn rather slowly, when at all, the "facts" of Euclidean geometry. Luneburg (1947) had proposed that visual space has a constant negative curvature, i.e. is Lobachevskyan. Battro *et. al.* (1976) have challenged this thesis on experimental grounds, holding instead that visual space has a variable curvature: it is sometimes positive (elliptical geometry), sometimes negative (hyperbolic geometry), and occasionally null (Euclidean geometry). As for the development of geometrical ideas, we know now that newborns move their heads to locate sounds and sights, so they do have some inborn spatial framework. But we also know, from Piaget's studies,

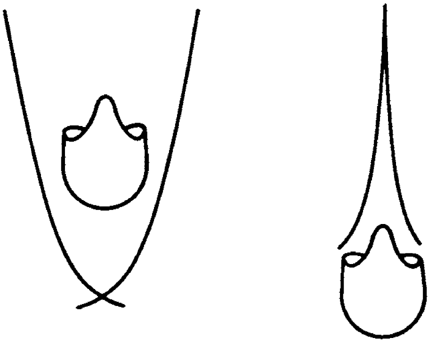


Fig. 4.4. Natural perspective is curvilinear. If asked to trace an avenue or parallel straight sides, the subject imagines curved lines. And if he looks at straight rail tracks he does not see them converging in straight lines but as branches of a hyperbola (Battro, 1979). Curvatures in drawing greatly exaggerated.

that this framework must be rather poor, for the triangle inequality does not appear normally before age 11, and even then only among half of the children (Piaget and Inhelder, 1967). Besides, our natural (untrained) perspective is not rectilinear but curvilinear (Battro, 1979): see Figure 4.4. In conclusion, we are not born Euclidean—or Riemannian or Lobachevskyan. If we did, geometry would not have been such a recent acquisition, and its teachers would have a far easier task.

3.2. *Appearance and Reality*

At birth we have only our perceptions to go by, so presumably we are born phenomenalists: only reflection upon our experience, together with education, turns us gradually into realists. At age 2 we may still reach for the Moon, but by age 4 we have already sobered up. The very young child presented with two things of the same volume but very different shapes, such as a ball and a pencil, tends to judge incorrectly their sizes: he sticks to appearances. Only the child who has attained the "concrete operational stage" tends to recognize the conservation of matter underneath changes in shape, color, and other appearances (Piaget, 1976a). The same holds for adult trained chimpanzees (Woodruff *et al.*, 1978).

We never see the world exactly as it is. For one thing we fail to perceive stimuli below a certain threshold and above a certain ceiling. For another we tend to inhibit most sensations. Thirdly, our bodily state changes from one moment to the next, so that a constant external stimulus is transduced now in one way, now in another. Fourthly, our neurosensors are such that "Humans do not experience the world as it is but as a power function of what it is" (Thompson, 1975, p. 239). This cryptic albeit true and profound statement may be understood as a verbalization of either Stevens' psychophysical law or Thouless' index. The former states that the perceived or apparent intensity of a sensory stimulus is a power function of the physical or objective stimulus. And the Thouless index relates the real and the perceived (apparent) sizes of a visual stimulus object, as well as the projection of the latter on the retina: see Figure 4.5. In sum, pure (sensory-motor) experience yields appearance not reality. If we wish to attain reality we must go beyond perception, into conception and action.

The perception of a fact is called a *phenomenon* or *appearance*. (In ordinary language 'phenomenon' is equated with 'fact': beware of the imprecisions of ordinary language.) There are imperceptible facts but there are no phenomena without sentient organisms. Appearance, then, is an

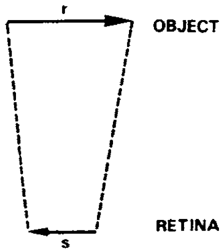


Fig. 4.5. The Thouless index: $i = \log(p/s)/\log(r/s)$, where p is the perceived or apparent size of the stimulus object of size r, whose projection on the retina is s. By algebraic manipulation we obtain $p = s(r/s)^i$. If $p = 4$, then $i = 1$; and if $p = s$, then $i = 0$.

evolutionary gain: it emerged together with the first animals equipped with nervous systems. Before them facts appeared to nobody: there was no appearance, there was only reality.

Phenomena are facts of a special kind, namely facts occurring in nervous systems. So, *phenomena are real.* Consequently there is no opposition between appearance and reality. My seeing the Moon larger on the horizon than overhead is a fact no less than the two positions of the Moon: only, the former is a perceptual, hence *subjective*, fact, whereas the latter are *objective* physical facts. There is then nothing wrong with admitting phenomena alongside nonphenomenal (or transphenomenal) facts. The opposition is not between appearance and reality but between subjective facts or accounts and objective facts or accounts.

From the fact that perception gives us only appearances, and the (false) hypothesis that perception is the only source of knowledge, some philosophers have concluded that there *are* only appearances, and others that only appearances can be *known.* The former (e.g. Berkeley) are *ontological phenomenalists,* the latter (e.g. Kant) are *epistemological phenomenalists.* Ontological phenomenalists hold that there is no world independent of the knowing subject: that to be is to perceive or to be perceived. Epistemological phenomenalists hold that, although there may be an autonomous world, i.e. a reality independent of the knowing subject, it cannot be known: at most we may construct fictitious worlds. Clearly, ontological phenomenalism entails epistemological phenomenalism.

A rigorous phenomenalist cannot go beyond appearances and must limit himself to describing or forecasting them: any attempt to understand them would lead him far beyond his own experiential horizon, into the wide world. Thus he will utter subject-centered sentences such as *Phenomenon x* *appears to subject y under circumstance z*. To the realist this is just a truncation of the full sentence *Phenomenon x, appearing to subject y under circumstance z, is caused by (or is indicative of) noumenon w*. And, in addition to such mixed sentences involving objective as well as subjective items, the realist will construct strictly objective sentences such as *Fact x occurs under circumstance z*. Indeed, he will attempt to translate all subjective and mixed sentences into strictly objective ones. This does not entail ignoring subjectivity but putting it in a wider context as well as trying to explain it.

Needless to say, scientists and technologists, regardless of their explicit profession of philosophical faith, behave like realists not phenomenalists. Thus if somebody complains that he feels *X*, his physician will probably conjecture that his patient has (objectively) ailment *Y*, which manifests itself as symptom (phenomenon) *X*. And the biologist will try to explain this cause-symptom relation in physiological terms. Nor is this realistic attitude confined to science and technology: it also permeates the arts. Thus Leonardo and Michelangelo studied anatomy in order to understand what they saw and thus be able to reproduce it more faithfully. Those who do not know "which nerve or muscle is the cause of each movement" are bound to "draw their nude figures looking like wood, devoid of grace" (Leonardo, 1492 or 1515, No. 488).

Phenomenalism, in either version, has been an obstacle to the inquiry into the real world, for it discourages the hypothesizing of objective facts behind phenomena, and it encourages instead the formation of superficial (phenomenological, behavioristic, black-box) hypotheses, and even the survival of an anthropomorphic world view. (Phenomenalism is the last ditch stand of anthropomorphism.) Nevertheless, phenomenalism has been embraced by a number of influential thinkers after Berkeley and Kant, such as Mill, Avenarius, Mach, Whitehead and at one time Russell, Carnap, Ayer, and Goodman. Besides, it has made a decisive contribution to the methodology of behaviorism (Watson, 1925), to the operationalist philosophy of physics (Bridgman, 1927), and to the so-called Copenhagen interpretation of quantum mechanics.

Phenomenalism, in either its ontological or its epistemological version, is false. First, as noted long ago by Plato (*Theaetetus* 158) against Protagoras, our ability to distinguish dreams and illusions from realities refutes the *esse est percipi* doctrine. Second, the very notion of a phenomenon or appearance is unintelligible apart from its dual, that of objective fact. (Thus it makes sense to estimate real sizes and relate them to apparent sizes only because we can distinguish them.) Third, every indirect observation, particularly if quantitative, employs hypotheses (such as law statements) that go beyond phenomena. Fourth, whereas some facts are observer-invariant, i.e. the same for all knowing subjects, others are not—and it is a task of scientific research to find out which are what. Fifth, the present can be explained only by assuming—as cosmology, geology, evolutionary biology and other sciences do—that there were real things before animals capable of perceiving them evolved. In short, the collection of all possible phenomena is only a (smallish) subset of the collection of all possible facts.

Yet, however insignificant appearances may be from an ontological point of view, they occupy a central position in epistemology. In fact, there is no way of gaining some deep knowledge about reality except by combining phenomena with hypotheses and processing both by reasoning. Not that, as James held (1890, II, p. 7), "Sensations are the stable rock, the terminus a quo and the terminus ad quem of thought"—for perception is fallible and has a very restricted scope. Phenomena are central to knowledge for different and various reasons. First because they *pose problems*—e.g. "Why do we perceive X?", "Why cannot we perceive X?", "Is X real or illusory?", and "Why does X appear to us now as Y, now as Z?" Second, because phenomena, when under control, *test* factual hypotheses—as in "If X is the case, then we must be able to observe Y. Now, we failed to observe Y. Therefore X cannot be the case." As James (1890, II, p. 301) put it, "What science means by 'verification' is no more than this, that no object of conception shall be believed which sooner or later has not some permanent and vivid object of sensation for its *term*. [...] Sensible objects are thus either our realities or the tests of our realities. Conceived objects must show sensible effects or else be disbelieved. [...] Strange mutual dependence this, in which the appearance needs the reality in order to exist, but the reality needs the appearance in order to be known!"

Since phenomena constitute only a very special kind of fact, real existence and the totality of real existents (i.e. reality) cannot be defined in terms of phenomena. In particular it is incorrect to define a real thing as a permanent possibility of being perceived (Mill, 1843, Bk. III, Ch. 24, Section 1). On the contrary, we must define appearance, or the totality of phenomena, as the collection of all (actual or possible) perceptual processes in all animals past, present and future. (We may also specify and speak of human appearance, blue jay appearance, sardine appearance, etc.) But, since we test existence hypotheses with the help of phenomena, the latter do provide a criterion of real existence. We propose the following *criterion* (not definition) of objective existence:

X exists only if the following conditions are met jointly:

(i) there are processes in X that produce, directly, or indirectly (i.e. via some other item), some perceptible effect Y;
(ii) this effect is observed by any competent observer in similar circumstances;
(iii) the link between X and Y can be rationally justified (preferably in terms of law statements).

Clause (i) includes the most commonly accepted reality criterion, namely "Being is acting (or having efficacy)". Since the universe as a whole cannot act upon anything, yet is real, this criterion, unlike ours, leaves out the universe. Clause (ii) is the condition of publicity or sociality of objective knowledge: an existence hypothesis proposed by a single individual, whether crackpot or genius, will not be accepted unless independently confirmed. And clause (iii) is the rationality condition: a dogmatic existence statement is unacceptable, for we must understand, at least in outline, the link between the phenomenon and the purported objective fact.

The above existence criterion is peculiar to *critical* (as opposed to naive) *realism*. This class of epistemological doctrines is contrasted in Table 4.1 to alternative views.

TABLE 4.1
Views on the relation of existence to perception.

Thesis	*Antithesis*
1. All perceptibles are real. (*Naive realism*)	1′. Some perceptibles are not (objectively) real.
2. All existents are perceptible. (*Naive phenomenalism*)	2′. Some existents are imperceptible.
3. All and only perceptibles are real. (*Phenomenalism*)	3′. Some existents are imperceptible or some perceptibles are not (objectively) real.
4. No existents are perceptible. (*Buddhism*)	4′. Some existents are perceptible.
5. All imperceptibles are real. (*Platonism*)	5′. Some (objects) are neither real nor perceptible.

4. CONCLUDING REMARKS

Only philosophers have questioned the principle that perception is necessary for knowing reality, and only philosophers have claimed that it is sufficient. Everyone else agrees that perception is necessary though insufficient: that it does yield knowledge (perceptual knowledge), though low grade and fallible. We all know that one and the same external object may appear small to an adult and big to a child, warm to an observer coming from cold outdoors and cold to another staying indoors. We all realize that we are not good at estimating directly vertical distances, very small or very large weights, and mental or social states. And many of us know also that by far the most numerous properties calculated or measured in factual science are just imperceptible. But we also know that any such properties, to be known, must be somehow related to sensible ones—via hypotheses, instruments, or both. True, perception is often deceptive: Figure 4.6. However, we can often discover and correct such perceptual errors, e.g. by checking with other senses, or other people, or by scientific analysis.

What holds for perception holds also for action and, indeed, for all types of experience: experience is no *solid* ground of knowledge but it is (metaphorically speaking) one of the two inputs to our cognitive apparatus—the other being reason. (Strictly speaking all experience is either past or future, hence hardly a solid basis for anything.) Strictly experiential judgments, of the forms 'I feel *X*', 'I do *Y*', 'I feel *X* when I do *Y*', 'I do *Y* when I feel *X*', and the like, are not objective because they report on a changing subject-object relation rather than on a subject-free object. Objective statements do not include reference to the subject. Such objectivity can be attained in either of two ways: really or illusorily. The

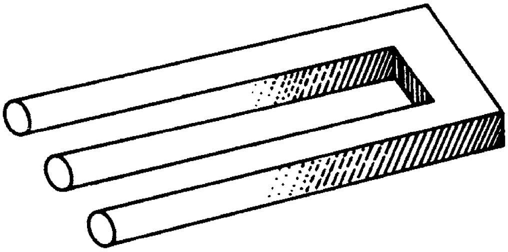


Fig. 4.6. The Devil's Trident: physically impossible. An illusion produced by trickery.

former by thinking up hypotheses about facts in themselves, i.e. independent of the way they are perceived, and putting such hypotheses to the test. This is the way of science and technology. The illusory way of attaining objectivity is the linguistic trick of substituting the impersonal 'It feels *X*', '*Y* is done', etc., for the personal 'I feel *X*', 'I do *Y*', etc.

Science and technology strive for objectivity, i.e. for knowledge of things in themselves. In other words science and technology discover or make facts but do not invent or construct them: what they do invent or construct is concepts, hypotheses, theories, methods, plans, and artifacts. This is not the view of subjectivistic philosophies, whether extreme like Berkeley's or moderate like Kant's. According to these philosophies there are no objective facts or, if there are any, they are unknowable: all facts would be man-made.

There are two kinds of subjectivism (or radical constructivism): individualist and collectivist. According to the former the individual knowing subject constructs the (his) world; according to collectivist constructivism, the world is an outcome of a collective activity. Individualist subjectivism is well known from classical philosophy, but it has hardly any followers among contemporary philosophers. Paradoxically it has some among theoretical physicists, particularly those who, following Bohr (1935), are wedded to the operationalist interpretation of quantum mechanics and state that nothing is an event ("phenomenon") unless it is observed by someone. Thus we are told that "We know now that the moon is demonstrably not there when nobody looks" (Mermin, 1981). (For a criticism of the subjectivistic misreading of quantum mechanics see Bunge (1973) and Vol. 6.) The following paradox demolishes this kind of subjectivism. If a subjectivist claims that nothing exists unless it is observed, then he must own that his own brain cells do not exist as long as their activity is not individually recorded. But this involves inserting an electrode in each of them, which, if technically feasible, would kill him. So, if the subjectivist wants to make sure that he is alive, he must allow a neuroscientist to kill him. It is better to give up subjectivism and stay alive.

Collectivist subjectivism is newer and less well known although it is rapidly gaining in influence. It was first formulated by Fleck (1935), adopted in a less extreme form by Kuhn (1962), and literally by a number of anthropologists and sociologists of science (e.g. Berger and Luckmann, 1966). Thus Latour and Woolgar (1979) claim to describe how Nobel laureate Roger Guillemin "constructed rather than discovered the fact of the thyrotropin-releasing factor". (Moreover he would have discovered such fact not as an outcome of observation, experimentation and hypothesis, but as a result of monitoring a number of "inscription devices" leading ultimately to the final inscription process called 'writing a paper', for which credit can be obtained. They do not explain why the whole process could not have been entrusted to a computer.) The only advantage of this type of constructivism is that it is self-destructive.

Empiricism held that perception is our point of departure, everything else being built upon it. (According to objectivist empiricism, such as Locke's, what we add is conception. On the other hand radical empiricism, such as Mach's, and the later James', claims that that world itself is made up of sensations.) Psychology has shown that there is no perceptual

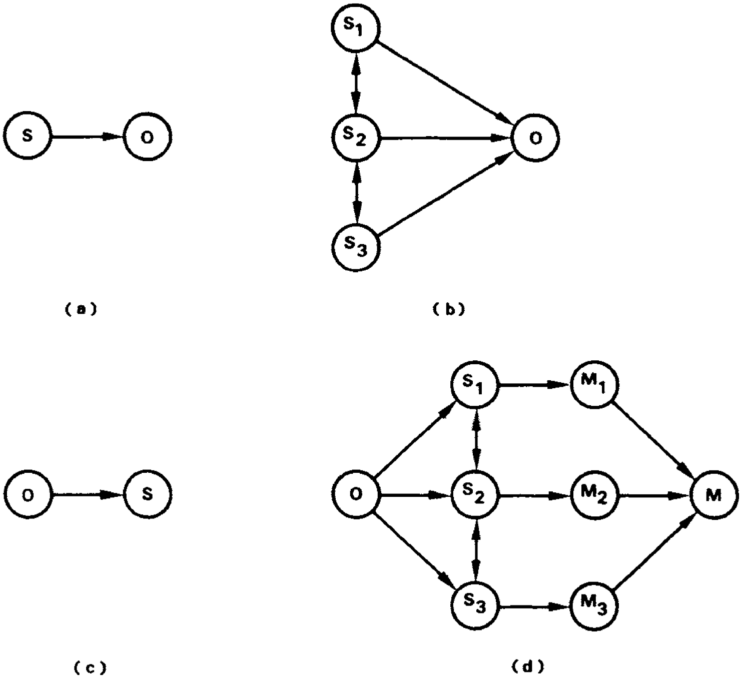


Fig. 4.7. Four views of the subject-object relationship. (*a*) *Individualist subjectivism*: the subject constructs the object, which is either a complex of perceptions (phenomenalism) or a theory (rationalism). (*b*) *Collectivist subjectivism*: the object is constructed by a community of subjects. (*c*) *Individualist realism*: the subject copies or mirrors the object. (*d*) *Social critical realism*: inquirers interacting in a community construct alternative models M_1, M_2, etc. of the object, which are tested until a model M is adopted by most of the members of the community.

"given" to be formulated as a set of indubitable basic sentences. We produce our own perceptions which are colored by our preconceptions, expectations, and social circumstances. All knowledge, whether perceptual or conceptual, is constructed—though not always deliberately.

What we are given is our external world and our genome. We do not have the burden of constructing the whole of reality: we are born into a self-existing world, which we manage to enrich in some respects and to mess up in others. And we are born with certain cognitive abilities that the natural and social environment can enhance or dim—in short, educate. Our perceptual limitations drive us not to constructing reality but to conceptualizing it. And the incompleteness and uncertainty of both individual perception and individual conception advise us to supplement them with, and check them against, those of similar inquirers. The result is that socially related knowing and acting subjects build fallible perceptual and conceptual representations of reality. This is the gist of our epistemology, which is a variety of realism, namely *biological* (as opposed to mentalistic), *social* (as opposed to individualistic), and *critical* (as opposed to naive and dogmatic). See Figure 4.7.

CHAPTER 5

CONCEIVING

We now come to ideas proper, in particular concepts and judgments. Other types of ideas, such as reasonings, problems, and rules, will be treated in subsequent chapters. Although some ideas originate in perception they all overreach it. Sit quietly in a sound-proof room, close your eyes, and think of any intellectual problem. You will be thinking up concepts, most likely not isolated concepts but entire systems of them.

We shall start by examining elementary ideas, i.e. concepts, such as those of book and equality. There are two different concepts of a concept: the psychological and the philosophical one. In psychology a concept is a mental (or brain) process; in philosophy it is a unit construct, or the smallest component of a proposition. In either case concepts are the building blocks of all other ideas. We relate the two concepts when we say that we are thinking of a given concept. Thought is a concrete process, hence one that takes time; on the other hand its object is abstract, in the sense that we detach it from such process and feign that it exists by itself as a timeless object. In this chapter we shall deal with the two notions of a concept, as well as with judgments and their philosophical counterpart, namely propositions.

1. CONCEPT

1.1. *From Percept to Concept*

When seeing or imagining a particular animal we say that we form a *percept* of it—or, more simply, that we *perceive* it. Such a percept, to have any cognitive value at all, must be in some respects analogous to the object perceived or imagined. On the other hand when thinking of humans in general—e.g. that they can conceive anything—we form *concepts* of them, e.g. those of *Homo sapiens, Homo faber, zoon politikon*, etc. Although a concept of this kind derives partly from our perceptions and memories of individual human beings, it is not identical with them. To begin with it is general, and so can bear no resemblance to any individual. A fortiori the so-called abstract concepts, such as those of transitivity, entropy, species, mind, or justice, are not similar to their referents even when they have them.

In other words, the perceptual representation of a thing is a (partial) function of its (apparent) traits into neural events, that preserves some of the relations among such traits (Ch. 4, Section 3.1). On the other hand a concept, or conceptual representation, of the same thing is no such map: it bears no resemblance to its referents, if any. Think of a theory of the electromagnetic field or of the competitive economy, each compressed into a system of equations. This does not mean that, unlike percepts, concepts are in an immaterial mind, let alone an impersonal Mind, rather than in some concrete brain. Concepts are brain processes (or collections of such) but they do not involve any neurosensors even though some of them do originate in perception.

In the above we have tacitly distinguished two kinds of concept: those which originate in perception and those which do not. We may call them *empirical* and *transempirical* (or transcendental) respectively. Idealists deny the existence of the former or render it dependent upon that of the latter. On the other hand empiricists and vulgar materialists (nominalists) deny the existence of transempirical concepts, and hold consequently that the words designating them are just meaningless sounds. Thus whenever Hume (1739) set out to evaluate an idea he asked for its psychological pedigree: *From what (sense) impression cou'd this idea be deriv'd*? Since he could not find the sensory roots of the metaphysical ideas of substance, cause, and self, he dismissed them. Likewise Kant (1787) professed to reject all the concepts of pure reason, particularly if they purported to refer to things in themselves: he too wished to stick to phenomena although he never succeeded. Needless to say, both Hume's and Kant's criticism of transempirical (in particular metaphysical) concepts holds only provided one accepts the empiricist and anthropocentric dogma that only that exists which can make some sense impression on us. (Recall Ch. 4, Section 3.2.)

If Kant and Hume had known the mathematics and physics of their time, they would have rejected them with the same vehemence with which they rejected metaphysics, for those sciences contained key transempirical concepts such as those of function, continuity, and infinity, speed of light, action at a distance, and mass. And they would have disapproved even more strongly of contemporary science, which involves a far more abstract mathematics, as well as factual concepts far removed from experience, such as those of field, electron, and gene. In particular, as Einstein noted, the concepts of theoretical physics are independent of sense experience and are not abstracted from it. "The relation [between concept and percept] is not analogous to that of soup and beef but rather of wardrobe number of overcoat" (1936, p. 353). Consequently "There is no inductive method which could lead to the fundamental concepts of physics" (*op. cit.*, p. 365).

To be sure the concepts of the empirical kind "derive" from percepts. However, they are not mere distillates or summaries of experience. The coining or learning of a new concept of either kind is a creative process: it consists in the emergence of something new that was not in perception, let alone in the external world. (From a neurophysiological point of view it consists in the emergence of a new plastic neuronal system: recall Ch. 1, Section 1.2. Such a process occurs in one of the higher brain centers. This explains why any empirical concept can be aroused by any one of the senses.) We may think of such a process as a map composition, or map of maps: from stimulus objects to percepts and from these to empirical concepts. See Figure 5.1.

The transempirical concepts do not originate in perception, i.e. they cannot be learned from experience but must be acquired by reflection. However, they are not necessarily isolated from all empirical concepts. For one thing some transempirical concepts may be suggested by empirical ones. Thus the empirical ideas of five fingers, five shells, and five children may have suggested forming the transempirical concept of number five. Conversely, some transempirical concepts may suggest empirical concepts and these in turn may guide some experiences. For example, thinking of low temperatures may suggest the idea or the image of a cold drink, which may in turn elicit the idea of paying a visit to the fridge.

In sum, although we emphasize the differences between percept and concept, and between empirical and transempirical concepts, we also assert that the items of these three kinds are not isolated but compose three-tiered *dynamical cognitive networks*. Each of them is related to some others, and each of them can modify some others. This view subsumes both empiricism and idealism, for it admits the creativity of both experience and the

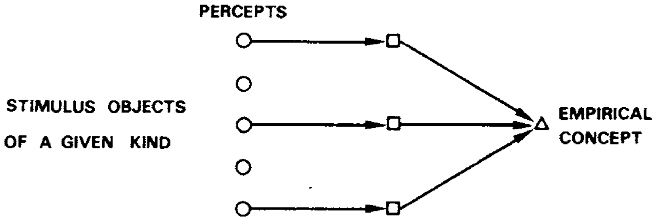


Fig. 5.1. The formation of empirical concepts as map composition.

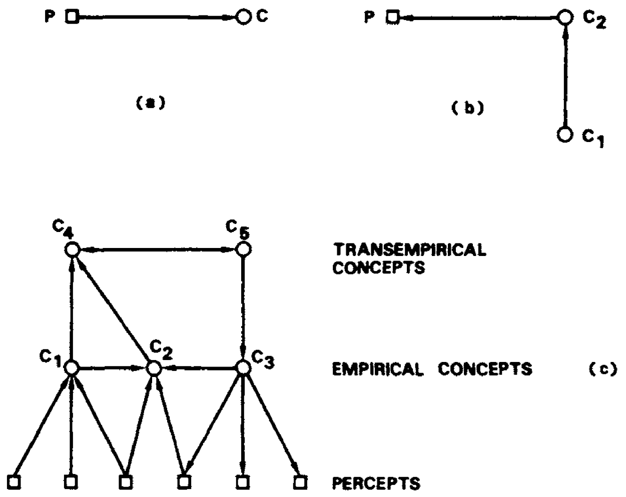


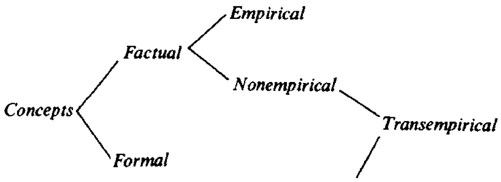
Fig. 5.2. Three views on the relations among percepts and concepts. (*a*) *Empiricism*: Every concept originates in a percept. (*b*) *Idealism*: Every concept originates other concepts, which in turn make some percepts possible. (*c*) *Realism*: Whereas some concepts originate in percepts, others do not; besides, some concepts guide perception and all concepts generate further concepts.

mind—or rather the brain. (See Figure 5.2.) But it transcends the borders of both classical epistemologies in that it suggests looking for the neuronal systems that do the perceiving, the empirical conceiving, and the transempirical conceiving as well as for the mutual actions among such systems. (Note that we are presupposing that perceiving and conceiving are activities of multineuron systems. The idea that there is one neuron for every concept lacks experimental support and is inconsistent with the existence of about 10^{11} neurons in the human brain. Indeed even a supergenius capable of learning one concept every minute would have accumulated, after a long lifetime, only about 50 million concepts, a tiny fraction of the total number of plastic neurons in the human brain: the vast majority would be idle.)

Among the transempirical concepts we distinguish the formal ones, i.e. those occurring in logic, mathematics, and semantics. These concepts are called 'formal' because they do not refer to any facts, whether objective or subjective (experiential). Thus the logical concept of conjunction concerns predicates and propositions, the mathematical concept of limit concerns functions and sequences, and the semantical concept of reference concerns predicates and propositions. That such concepts have no factual reference can only be proved with the help of some theory of reference. (Cf. Vol. 1, Ch. 3.) That, in spite of such lack of factual reference, they can be "applied" in factual science and technology, is another matter to be discussed in Vol. 6.

The admission of formal concepts was a major sacrifice for contemporary empiricism (or logical positivism): it had to be made in order to use logic to analyze and organize empirical knowledge. On the other hand the materialist suffers no setback in admitting formal concepts: all he needs is the hypothesis that all concepts, even the formal ones, are brain processes or collections of such (Bunge, 1981a).

According to logical positivism all facts are experiential, so that all concepts with a factual reference must have an empirical content or meaning, i.e. must be empirical concepts. Hence according to that philosophy the factual/formal dichotomy coincides with the empirical/formal one. This classing is wrong because it is not exhaustive. Indeed there are nonexperiential facts, such as atomic collisions and species extinctions, and we have formed the corresponding transempirical concepts. In other words, whereas some factual concepts are empirical (e.g. "hot") others are transempirical (e.g. "temperature"). This is a firm result of a semantical analysis of scientific concepts (Bunge, 1973, 1974a, 1974b). The concept classification that accounts for the (conceptual) existence of nonempirical factual concepts is this:



Are there any laws of the development of concepts, or rather of our conceptual abilities? We know next to nothing about the conceptual development of children before the first year of age, and this mainly due to their very poor motor and communication abilities. However, neuroanatomical studies show that the cortex of the infant brain is poorly organized and therefore unlikely to perform any conceptual feats: it has small neurons, incompletely myelinated axons, a small dendritic arborization, and few synaptic connections. Older children have been investigated by Piaget and his school for over half a century. These studies have found that there are indeed definite stages in development that cannot be skipped. In particular they have found that the young child has hardly any general concepts: he or she acquires them in the course of development—the pace of which is presumably accelerated by formal education. Once the child has acquired some general concepts—such as that of causation—and some general propositions—such as the principle of conservation of matter—he can use them to construct further concepts as well as to face experience in novel ways.

As for conceptual evolution, we know even less. However, it is likely that all hominids, and perhaps many other higher vertebrates as well, have formed certain general concepts, such as those of existence, presence, place, time, notion, lightness, hotness, hardness, nearness, usefulness, and friendship. A methodical study *à la* Piaget of the conceptual abilities of apes, dolphins, ravens and other animals usually rated as intelligent, is overdue. Also, it would be interesting to investigate the tacit knowledge that must have gone into the fashioning of artifacts by primitive and ancient man. We know little about this in part because of the epistemological prejudice that every concept is preceded by some percept or some action. We know much more about the history of ideas. One thing we have learned from it is that, at least from the beginning of civilization, it is just not true that every complex idea has been preceded by simpler ideas. Whereas much progress consists in complication, some of it involves simplification. Besides, some very basic ideas, such as those of set and concatenation, were introduced long after far more complicated ideas—e.g. those of the infinitesimal calculus—had been invented. In short, there are laws of conceptual development but not of conceptual evolution.

1.2. *Concept Formation*

Let us now examine a handful of very basic conceptual operations. They are so basic that we usually take them for granted and therefore do not pause to analyze them. After all, it behoves philosophers to puzzle what others take for obvious.

The most basic relation among any two items, whether factual or conceptual, is that of difference. The perception or conception of difference is called *discrimination* (or *analysis*). Its dual, the failure to differentiate, is of course confusion. Our perceptual apparatus has a very restricted discriminatory (or analytic or resolving) power, but we can enhance it with the help of observation instruments. Likewise we can sharpen conceptual discrimination or analysis with the help of logic and mathematics.

We must guard against the temptation of detaching whatever is distinguishable. For one thing, properties and changes thereof (i.e. events) are not detachable from the concrete things that possess the properties or undergo the changes. (See Vol. 3, Ch. 2.) For another, every thing is a system or a component of some system, and every system but the entire world interacts with some other system. (See Vol. 4, Ch. 6.) The components of a system, though distinct and therefore distinguishable (at least in principle), are not separable. (Consequently it is impossible to specify the state of one of them without reference to the state of the others. I.e. the state space of the whole does not equal the union of the state spaces of the components.)

All of the mistakes alluded to above have been erected into philosophical doctrines. Thus (philosophical) *atomism* conceives the world as a heap (not a system) of detached or separate units (atoms): it has no use for systems possessing (emergent) properties not possessed by their components. Its dual, namely *holism*, prefers confusion: it proclaims that whatever cannot be detached is one. (See Vol. 4, Ch. 1.) Holists abhor analysis believing that it destroys wholeness, while actually it is the only way to understand wholes. Finally *dialectics* declares differents to be opposites, while at the same time proclaiming the unity of opposites. Since it does not analyze the concepts of opposition and of unity, dialectics is not easy to understand, hence to evaluate. (See Bunge (1981a) for a criticism.)

Discriminating between two objects is a modest cognitive achievement—even though, as in astronomy, it may call for an enormous cognitive investment. The next step is to find out what the distinct objects differ in. And this involves taking cognizance of some of their properties. Such cognitive operation is no mere perception: it is a conceptual operation, for it consists in attributing properties. Indeed, in order to be able to attribute a property to an object we must form some concept of such property. This concept may precede the act of attribution or it may be improvised during it, but in any case it must be formed in order for predication to occur. A concept of a property is called an *attribute* or *predicate*. (There are of course unary, binary, and in general *n*-ary predicates. Thus the predicate *between* is ternary, and so is any function of two variables.)

Idealists need not adopt our distinction between properties and predicates, but realists must, and this for the following reasons. First, a property of a thing cannot be detached from the latter: there are no (substantial) properties without things, and no things without properties

(Vol. 3, Ch. 2). On the other hand a predicate, if attributed (truthfully or falsely) to a concrete thing, is a conceptual representation of a thing's property. Second, attribution is a psychological process, hence one studied by psychologists, whereas all sciences are supposed to study properties objectively possessed by objects. Third, one and the same property can often be conceptualized (represented conceptually) by more than one predicate. Thus on a first level of analysis heaviness is representable as a unary predicate ("body *b* is heavy"); on a second level as a binary predicate ("body *b* in gravitational field *c* is heavy"); on a third level as a ternary predicate ("the heaviness of body *b* in gravitational field *c* equals *d*"), and so on. (In general, predicate formation can be construed as a function from properties to sets of predicates.) Fourth, whereas for every attribute there is another attribute equal to the negation of the first, things have only "positive" properties. A thing either possesses *P* or does not possess *P*, but it cannot "possess" not-*P*: negation is *de dicto* not *de re*. (Barnacles do not think, but this is not to say that they exert the function of not thinking. Negation affects the proposition "Barnacles think", not the property of thinking.) Fifth, what holds for negation holds also for disjunction: there are no disjunctive properties although there are disjunctive predicates. There is no such thing as the property of being alive *or* dead, although the predicate "is alive or dead" is perfectly respectable, and moreover tautological. In short, whereas the set of predicates of a given rank and a given reference class forms a Boolean algebra, the corresponding set of properties does not. (See Vol. 3, Ch. 2.)

There are two different ways of forming the concept of a property, i.e. of construing a predicate: intensionally or extensionally. In the former construal the predicate is a function from individual objects to propositions (Vol. 1, Ch. 1, Section 1.2). For example, "Metabolizes" assigns to organism *b* the proposition "*b* metabolizes", and "Interacts" assigns to the ordered pair of things *a* and *b* the proposition "*a* interacts with *b*". The truth value of such propositions is immaterial for the construction of the predicate. However, once we have found out such truth values we can form the extension of the predicate, i.e. the collection of individuals (or pairs, or *n*-tuples) for which the predicate holds. And then we can ascertain whether a given individual (or *n*-tuple) belongs to the extension of a certain predicate. In other words, the ordinary language expression '*b* is *F*' can be construed either as '*Fb*', where *F* is a function, or as '*b* is a member of the extension of *F*'.

Ordinarily the extensional construal presupposes the intensional one, for we must know what predicate we are talking about (and what property it conceptualizes) before inquiring into its extension. However, both construals are equivalent, and mathematicians use them interchangeably. (For example, the axiom of mathematical induction, occurring in arithmetic, can be formulated in either of the following ways: (i) If a property holds for zero and, when holding for an arbitrary number, it also holds for its successor, then it holds for every natural number: (ii) If a set of natural numbers contains zero, and contains also the successor of every natural number it contains, then it contains all natural numbers.)

Nominalists trust only individuals and mistrust concepts and therefore predicates. To them every property must be understood as the collection of individuals possessing it: they conflate predicates with their extensions. This confusion is open to several objections. The most damaging ones are as follows. First, let P_1 and P_2 be two properties of entities of a certain kind K, e.g. the mass and momentum of bodies. Since all K's possess both P_1 and P_2, according to nominalism P_1 is identical to P_2—which contradicts the hypothesis that P_1 and P_2 are different. Besides incurring contradiction, extensionalists and nominalists are confronted with a practical difficulty: they must know all the values of a given function in order to be able to speak of the function. But in practice this is seldom the case: we are not given complete tables but differential equations, recursion formulas, or some other condition defining the function. Moreover we must know such defining conditions in order to be able to compute (some) values of the function. (In addition we need some special technique, such as series expansion, or iteration, or numerical interpolation.) In short, we cannot sacrifice science and mathematics to philosophical prejudice.

When forming concepts we should avoid the dual sins of reification and ideafication. *Reification* is of course the incorrect conception of properties, relations, or even concepts, as entities having an autonomous existence. Classical examples are the idea that sickness is an entity that the patient carries and may pass on to someone else, and the medieval conception of darkness as an entity (*tenebras*). More recent examples are the existentialist notion of nothingness as a thing, and the structuralist idea that structures precede the corresponding structured things. The dual procedure is *ideafication*, or the conceiving of material things or processes as self-existing ideas. Classical examples are the idealist conception of ideas as detached from thinking brains, and of culture as a system of ideas, values, norms, and behavior patterns. To be sure it is permissible, nay sometimes indispensable, to *feign* that mental processes are detached from brains, in order to be able to focus on certain common features of concepts and disregard everything else, in particular the thinkers' circumstances. This is merely an instance of what may be called *methodological abstraction*. But we should keep in mind (or rather brain) that such conception of concepts in themselves is a fiction.

In studying properties we shall do well to distinguish three kinds of them: (*a*) *individual* or possessed by individuals (e.g. the mass of a molecule, the demand for a certain commodity by a certain individual consumer); (*b*) *aggregate* or possessed by aggregates (e.g. the total mass of a conglomerate of bodies, the total demand for a certain merchandise by a society); (*c*) *emergent*, or possessed by a system but not by its components (e.g. the stability of a molecule, the type of an economy). Failure to draw these distinctions results in absurds such as "the average molecule" or "the average consumer" (instead of "the average of property P over an aggregate of molecules—or consumers"). Two cautions are in order. The first is that a number of properties that can be treated as aggregative (hence additive) to a first approximation are actually systemic. This is the case with mass and entropy, as well as with aggregate demand and supply in economics. (Consider: if two blue collar families are friends and neighbors, they can pool their resources to buy one car, whereas separately neither of them might afford to.) Second, a number of properties that seem to be possessed by individuals are actually possessed by components of a system and vanish if the system breaks down. (For example, the property of being a faithful spouse vanishes upon divorce. Likewise some of the properties of producers and consumers depend on the type of economy.)

There are essentially three ways of studying properties: (*a*) to study *particular* properties, such as the surface tension of a given liquid or the feeding habits of animals of a given species in a given environment; (*b*) to study the *distribution* of a given property in a population, e.g. the distribution of fertility in an animal population, or of wealth in a human society; (*c*) to study the *relations between two or more properties* of a thing, e.g. the correlation between unemployment and crime. All three are legitimate concerns, so neither should exclude the other, and all three should be carried out both theoretically and empirically.

Perceiving or conceiving commonalities is no less important, for attaining knowledge, than realizing differences. Two items are said to be *equivalent* (or *equal*) in some respect, i.e. with respect to a given property, just in case they share this property. This definition entails that there are as many equivalence relations as there are properties. It also entails that every equivalence relation is reflexive, symmetrical, and transitive. I.e., for any objects a, b and c in the given set: $a \sim a$ (reflexivity), if $a \sim b$ then $b \sim a$ (symmetry), and if $a \sim b$ and $b \sim c$, then $a \sim c$ (transitivity).

If the transitivity condition is relaxed we obtain the weaker relation of *similarity* or *resemblance*. If a is similar to b, and b similar to c, it may or may not be the case that a resembles c. That is, a and b may share some properties which are not those that b and c share: see Figure 5.3. A qualitative measure of similarity is the intersection of the sets of properties of the similar objects. That is, calling $P(a)$ and $P(b)$ the sets of properties of a and b respectively, we stipulate that the similarity $s(a, b)$ between a and b equals $P(a) \cap P(b)$. By counting the number of shared properties we obtain a quantitative measure. Our qualitative measure allows us to introduce two relations of similarity comparison:

$$a \textit{ is more similar to } b \textit{ than to } c =_{df} s(a, b) \supseteq s(a, c),$$

$$a \textit{ is more similar to } b \textit{ than } c \textit{ to } d =_{df} s(a, b) \supseteq s(c, d).$$

If two items are distinct (i.e. really two instead of one), and yet equivalent in some respect, then they can be put into a single class. (Not so if they are merely similar.) Note the expression 'can be put' instead of 'exist'. Differences among factual items are objective: they exist whether or not a given subject knows it. On the other hand grouping is a conceptual operation: classes are concepts, not real (concrete, material) entities. But of course whereas some groupings are arbitrary or artificial others are natural or objective. Thus by putting together all the persons called John we obtain an artificial class, whereas by grouping all the persons sharing some common ancestors we obtain a natural class, i.e. a collection whose members are objectively related. (For the notion of natural class see Vol. 3, Ch. 3.)

The operation of grouping or collecting is a basic animal behavior: all animals gather things of some kind or other, and the higher vertebrates are also capable of collecting ideas, i.e. of grouping images, concepts, propositions, and the like. Material things can be put together spatially or

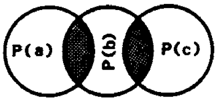


Fig. 5.3. The properties shared by a and b are not the same as those b and c have in common. Hence, although a is similar to b, and b similar to c, a does not resemble c.

temporally; and all objects, regardless of their nature, can be put together in thought. In the former case the outcome is a material thing: a conglomerate or a system; in the second case the outcome is a concept: a set. And in either case the outcome is assumed to be independent of the order in which the items have been gathered: i.e. the result is either an unordered aggregate or an unordered set—or perhaps the matter of ordering has been disregarded on purpose. We are concerned here with collecting items of any nature whatever, whether factual or conceptual, to form conceptual collections, i.e. classes. Whereas some classes are realistic (not arbitrary), no class is real. We shall come back to this subject in relation to biological taxonomy (Vol. 6).

(The operation of collecting or grouping can be defined as follows. Consider three items: 0, a, and b, forming the following binary compounds or concatenates: $00, 0a, a0, 0b, b0, aa, bb, ab$, and ba. Assume that $00 = 0$, $0a = a0 = a$. $0b = b0 = b$, $aa = a$, $bb = b$, and $ab = ba$. In other words, 0 is the null object, and concatenation is idempotent and commutative. We now define the *collecting* or *bracketing* operation $\{\ \}$ by the following assignments:

$$0 \mapsto \{\ \};$$

$$a \mapsto \{a\}, \qquad b \mapsto \{b\};$$

$$ab \mapsto \{a, b\}.$$

From the assumed properties of the concatenation it follows that

$$aa \mapsto \{a\}, \quad bb \mapsto \{b\}, \quad ba \mapsto \{a, b\}.$$

In the above, a and b may be any objects, whether material or conceptual. On the other hand 0, the null object, can only be conceptual: nothingness has no real existence. Finally, we define $\varnothing =_{df} \{\ \}$.)

When collecting real (material) things we must recall that they are changeable, to the point that whereas some of them may have already vanished others may not yet have come into being. (Think of mankind, or the human species.) We must therefore distinguish between a collection that varies in time and a set, or collection regarded *sub specie aeternitatis*. The membership of the former is variable, whereas that of the latter is constant: all the elements in the latter are treated as if they were timeless entities. The difference between the two collecting operations can be elucidated with the help of the notion of time, which is of course a scientific and ontological concept, not one in pure mathematics. In fact call F the attribute of interest and write 'Fxt' to indicate that x has property F at time t. Then the *collection of all F's at time t* will be

$$S_t = \{x \mid Fxt\}.$$

(Instead of augmenting the n-arity of predicates we can also use the operation of bracketing at a given instant or over a given time interval, or $\{\ \}_t$ for short. This operation can be defined by adding t to the individuals $0, a$, and b used to define the timeless bracketing operation $\{\ \}$, and requiring that it obey the following conditions:

$$0t \mapsto \{\ \}_t;$$

$$at \mapsto \{a\}_t, \qquad bt \mapsto \{b\}_t;$$

$$abt \mapsto \{a, b\}_t.$$

The expression '$\{a, b\}_t$' is read 'the set formed by a and b at (or during) t'.)

By taking the union of all the variable collections of F's for all time we get the (timeless) *set* of Fs, or the *extension* of the predicate F. (In physics many predicates are frame dependent, so it may be necessary to add a subscript naming the particular reference frame relative to which the collecting is done.)

Just as we can analyze individual things and individual ideas, so we can analyze collections of either. The most basic mode of analysis of a collection is its *partition* into homogeneous subcollections. For example, a human population can be analyzed into collections of individuals of the same sex, the same age, or the same number of schooling years, etc. The key to such partition is the concept of sameness in some respect, i.e. of equivalence, examined a while ago. One says that an equivalence relation *induces the partition* of a collection into a family of *equivalence classes*, and that all the members of each equivalence class are equivalent (or equal) in the given respect. The operation resulting in the formation of such family is called the *quotient* of the given set S by the given equivalence relation $\sim$ (written '$S/\sim$'). For example, the quotient of the set of integers by the relation of equal parity is the family composed by the set of even numbers and the set of odd numbers. (I.e. $\mathbb{Z} = \{E, \bar{E}\}$, where $E = \{x \in \mathbb{Z} \mid (\exists y)(x = 2y)\}$, and $\bar{E} = \{u \in \mathbb{Z} \mid (\exists v)(u = 2v + 1)\}$.) We shall use this notion when studying the principles of classification (Ch. 9 Section 1.1).

It has been noted that, whereas some people excel at analyzing, others are better at synthesizing: few are at the same time good splitters and good lumpers. Yet both abilities are needed, for whereas some lines of inquiry start by distinguishing or resolving, others begin by fusing or lumping, and all research projects include these two moments. Indeed, analysis does not go far without interrelation, and in turn interrelation may result in the recognition of wholes. So, splitters as well as lumpers discharge a useful function in any inquiring community—provided each of them does not enshrine his own ability as a philosophy (atomism and holism respectively) and proclaims the superiority of it.

So far we have encountered concepts of three basic types: individuals, collections of such, and functions. In the present stage of the evolution of ideas every concept is either an individual concept (e.g. "Earth"), a collection (e.g. that of all solar planets), a function (e.g. "not"), or some construction out of concepts of such basic types. For example, an ordered n-tuple may (but need not) be construed as a certain set of sets. (E.g. $\langle a, b\rangle = \{\{a\}, \{a, b\}\}$.) A vector may be analyzed as an n-tuple of real numbers and also as a unit component of a vector space.

The general concept of an n-tuple and the more special one of a vector are formal concepts. However, they can be applied to real life situations. Consider a set of commodities (goods or services). Call q_i the quantity of the ith commodity, measured in appropriate units, and p_i the corresponding unit price in dollars. Then the value of the amount q_i of commodity i is by definition the product $p_i q_i$. A collection or bundle of commodities q_1, $q_2, \ldots, q_n$ can be conceptualized as an ordered n-tuple. Prices are parallel. The price structure or value of a bundle of commodities is then defined as the product $v(q) = pq$ of the vectors $p = \langle p_1, p_2, \ldots, p_n\rangle$ and $q = \langle q_1, q_2, \ldots, q_n\rangle$. This notion allows one to compare the values of different commodity bundles. In fact we stipulate that $v(q) \geq v(q')$ if and only if $pq \geq p'q'$. In this manner we can conceptualize a whole bundle of commodities as a single concept, and reduce the comparison of two bundles of commodities to comparing two numbers. Moreover all of the concepts involved in the above process are exact.

The general concept of an n-tuple is abstract, that of an n-tuple of real numbers (or vector) is less so, and that of a commodity bundle is not at all abstract. This kind of abstraction may be called *semantical* because it is the dual of semantic interpretation. In fact we get the notion of a vector by interpreting the components of an n-tuple as real numbers, and the notion of a commodity bundle by interpreting in turn those real numbers as quantities of commodities. In other words the factual concept (of commodity bundle) results from the factual interpretation of a mathematical interpretation of an abstract concept. (For the composition of the two interpretation maps see Vol. 2, Ch. 6.) Semantical abstraction is, in short, the dual or inverse of interpretation.

Another concept of abstraction may be called *methodological*, as it consists in deliberately disregarding certain properties. It is identical with that of equivalence. Indeed, if in comparing two objects we abstract from (disregard) all their properties except those they share, we build an equivalence relation allowing us to partition the given set into equivalence classes. (Recall what was said about equivalence a while ago.) An instance of this notion of abstraction is the formation of the concept of relation apart from the relata. Thus take all the possible statements of the form "x lies to the right of y", and abstract from the relata, i.e. consider the relation of lying to the right: this concept is methodologically abstract. Clearly, every time we generalize and every time we build a model of a real thing we disregard certain properties and thus perform a methodological abstraction.

A third concept of abstraction is *epistemological*, for it consists in remoteness from ordinary experience, in particular perception. Epistemological abstraction consists in replacing phenomenal properties with nonphenomenal ones, hence in replacing empirical concepts with transempirical ones. (Recall Section 1.1.) For example, the empirical (subjective) statement "I feel colder now" suggests either of the following transempirical (objective) statements: "The air temperature has dropped" or "There may be something wrong with the subject's health". All mathematical concepts and most of the theoretical concepts in science and technology are epistemologically abstract. Only psychology employs phenomenal concepts, such as that of loudness, alongside transempirical ones, such as sound intensity— and this for the good reason that it wishes to account for subjective experience, not only for objective facts.

The three concepts of abstraction are different but they are logically related. The relation is this: Whatever is either semantically or methodologically abstract is also epistemologically abstract—but the converse is false. All three concepts seem to come in degrees. However, so far only semantical abstraction has been quantitated (Vol. 2, Ch. 6). The quantitation of methodological and epistemological abstraction is an open problem. But at least in a few cases we can order concepts according to epistemological or methodological abstraction. For example, the notions of electromagnetic potential, electromagnetic field intensity, and ponderomotive force, form a ladder of descending epistemological abstraction.

And the concepts of vertebrate, animal, and organism form a ladder of ascending methodological abstraction.

Finally, are there inconceivables, i.e. facts that cannot be conceived of, or concepts that cannot be formed? If there are any we cannot say a priori which they are for, if we did, we would have some conception of them. What we may freely grant is that some facts will forever remain unknown to us, and some conceptual processes are beyond our capacity. But we cannot always specify which ones. Still, we should remember that a number of inconceivables have turned out to be conceivable. Two classical examples are the irrational numbers and change. The former were inconceivable within the Pythagorean doctrine, which held that every number was either an integer or a ratio of integers. As for change, from Parmenides to Bergson a long string of philosophers proclaimed that, although it may be intuited, it cannot be conceived. Yet modern mathematics enlarged the narrow Pythagorean context and moreover proved that there are infinitely many more irrationals than rationals. And modern science places equations of change (motion, field propagation, chemical or nuclear reaction, organic growth, social mobility, etc.) at its very center. So, it is more prudent to talk about *unconceived for the time being* than about inconceivables. And in any case it is more fruitful to try and form new concepts than to speculate about our unknown conceptual limitations.

2. PROPOSITION

2.1. *From Thought to Proposition*

We are now ready to examine complex conceptual operations and, in the first place, thoughts such as thinking that psychology is interesting, that psychology must be allied to sociology, and that all conceptual operations are neurophysiological processes. We have just exemplified three modes of judgment: attribution, association, and generalization. We shall study these basic modes of judgment after the following preliminaries.

Thinking is a brain process the psychological units or modules of which are concepts. Each concept is identical with the specific activity or function of some psychon or plastic neuronal system. (Ch. 2, Section 2.2.) Forming a judgment is thinking of several concepts in quick succession or simultaneously. Therefore judging is identical with the sequential activation of a number of psychons—or perhaps with the simultaneous activation of them. And believing is even more complex, for it is identical to

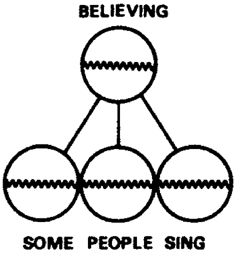


Fig. 5.4. Schematic representation of the sequential activation of the psychons constituting the judgment that some people sing. The evaluating psychon on top does the believing (or disbelieving).

giving assent to such thoughts. Hence presumably believing involves not only conceptual psychons but evaluating ones as well. See Figure 5.4.

Every judgment is a sequence of concepts but the converse is false. Thus the sequence of concepts expressed by the phrase 'She got' is only part of a judgment. We know how to characterize a complete sentence and a well formed proposition, but we still do not know what characterizes a complete judgment. That is, we still have to learn what are the neurophysiological characteristics of a complete judgment as distinct from an arbitrary sequence (or association) of active psychons. This will force us to work most of the time with the conceptual and linguistic manifestations of judgments, i.e. with propositions and sentences respectively.

We may as well start by emphasizing the difference between judgments and propositions. Whereas the former are brain processes, the latter are collections of such and, more particularly, equivalence classes of judgments (Ch. 2, Section 2.2). Consider an external fact such as the song of a bird in a neighborhood. A number of observers may think of this fact, and some of them may agree that it is too early for this time of the year. Most likely the brain processes the various observers go through when thinking this are, though equivalent in some respect, different in other respects. Yet they must be the same in one respect, namely that all of them consist in thinking that the perceived song is too early for this time of the year. And this is precisely the proposition designated by the English sentence 'That song is too early for this time of the year', the Spanish sentence 'Ese canto es muy temprano dada la estación del año', and hundreds of others. Moreover each of these sentences can be uttered in a variety of ways, and likewise each of those judgments can be believed with different intensities. Figure 5.5 summarizes these distinctions, which unfortunately are not always made.

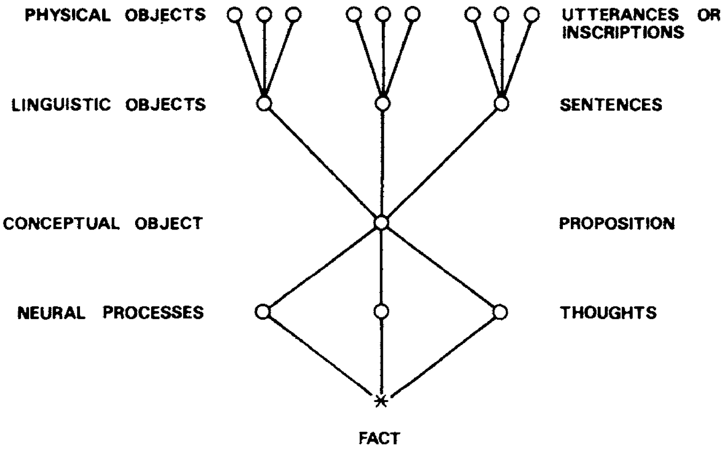


Fig. 5.5. The proposition as an equivalence class of thoughts and the sentence as an equivalence class of utterances or inscriptions.

Some of the categories we have just mentioned are deliberately fused for philosophical reasons. (Fusing is not the same as confusing: the former is conscious equation, the latter is lack of discrimination.) Thus propositions are sometimes called 'facts', perhaps under the influence of idealism. And it is still common to call propositions 'sentences' (and accordingly to speak about the 'sentential calculus') because of the nominalistic prejudice against conceptual objects. Although in most cases such confusions are harmless, in others they are misleading.

(An example of the latter is the problem area of counterfactuals. An entire industry has been built upon the mistaking of counterfactual conditionals for propositions, while actually they are not even sentences. In fact, counterfactuals are ambiguous utterances or inscriptions. Consider the time-worn counterfactual 'If this lump of sugar were placed in water it would dissolve'. This expression may be interpreted in at least two different ways: either as emphasizing that the given lump of sugar is *not* actually immersed in water, or as emphasizing that, when placed in water, it *dissolves*. The reading is a matter of stress or of italics. In the first case we would say 'If this lump of sugar *were* placed in water it would dissolve'. In the second, 'If this lump of sugar were placed in water it would *dissolve*.' Without some information about the context, in particular the speaker's or writer's intention, it is impossible to disambiguate the utterance or the inscription. Because of this ambiguity counterfactual conditionals have no logical status and do not occur in mathematical or scientific theories. They constitute a pseudoproblem. (See Bunge, 1973b, Ch. 1.) No wonder it is being approached by the bizarre metaphysics of possible worlds, which is not a theory about the only world there is (*Treatise*, Vol. 3, Ch. 4).)

Let us now perform a quick study of the operations of attribution, association, and generalization. *Attribution* is the conceptual operation of assigning a property or a relation to objects of some kind, material or conceptual. Thus when perceiving that *b* is moving we may attribute to *b* the property of moving—or, on second thoughts, we may attribute to ourselves the property of moving relative to *b*. And on perceiving *b* crashing against *c* we may attribute to them the relational property of colliding with one another, or even that of attracting each other. (One and the same percept may elicit no judgment, one judgment, two judgments, or even more.)

When attributing phenomenal (observable) properties to perceptible individual things we may not distinguish between perceiving and conceiving, and may even be tempted to espouse phenomenalism. The difference emerges when thinking of properties regardless of their bearers (e.g. "swiftness") or when attributing nonphenomenal properties (e.g. "external"). In either case we obviously go beyond perception and form concepts: in the first case we generalize, in the second we introduce transempirical concepts.

We may conjecture that attributing property *P* to object *b* consists in the (simultaneous or successive) activation of the psychons for *P* and for *b*. Or we may conjecture that we first think that *b* is (possesses the property) *P*, with a single psychon, and then the psychons for *b* and *P* are activated (or formed) at the same time or sequentially. Or we may also conjecture that sometimes we undergo processes of the first type (synthesis) and at other times processes of the second type (analysis). We still do not know which, if any, of the three hypotheses is true, and we won't find out as long as they are not spelled out (if possible mathematically) and tested experimentally. However, some crude conjectures have to be hazarded before they can be rendered more precise and put to the test.

Things are easier at the conceptual and linguistic levels, where it is a question of forming propositions or sentences, i.e. of gluing predicates to subjects. Consider a predicate or attribute *F* representing (conceptualizing) an intrinsic or non-relational property *P* of objects of some kind *K*. *F* can be construed as a function from *K* to the set of all possible propositions involving *F*. And attributing *P* to a member *b* of *K* is forming the proposition *Fb*, i.e. evaluating the function *F* at *b*.

Association, in particular pairing, is the general case of attribution. Associations can be natural, as those established between lightning and thunder, or hunger and food; or they can be conventional, like those between side of the road and sense of circulation, or house and street number. And objects of any kind may be thought (truly or falsely) to be associated: things to things, events to events, things to events, concepts to concepts, concepts to properties, etc. Whatever its nature and origin, thought of association may be conjectured to be a neural process similar to that of attribution, i.e. either the pairing of psychons or their sequential activation. The mathematical notion of a relation, or n-ary predicate, elucidates that of association at the conceptual level. Thus consider a relational property P of objects of kinds K and L, such as the relations of connecting, eating, or cooperating. P can be construed as a binary predicate G with domain the cartesian product of K and L, and codomain the set of all propositions containing G. In this case attributing P to the pair of objects b in K and c in L consists in forming the proposition Gbc, i.e. in evaluating G at $\langle b,c \rangle$.

Finally we come to *generalization*. There are two kinds of generalization: (*a*) from like cases, i.e. from individuals to species; (*b*) from one species to another. The former consists in discovering or conjecturing a pattern from individual cases. The second type of generalization is what learning theorists call *transfer of learning* (or of skills)—e.g. having learned one foreign language facilitates the learning of another, or learning carpentry facilitates the learning of work with metals. We shall limit ourselves to generalization of the first type.

How do we generalize? Empiricists hold that every general proposition, in particular every law-statement, is an inductive generalization, i.e. a drawing together of particular propositions. Thus every general hypothesis would come at the end of an inductive process, never at the beginning of a testing procedure. On the other hand rationalists and intuitionists hold that we can start with generalities straight away, on the strength of a single case or even a priori. They also believe that no tests are necessary. Experimental psychologists have found that in fact we and other higher animals proceed sometimes as empiricists and other times as rationalists. Although some generalizations culminate repeated observations, others are formed on the strength of a single case or even prior to observation. That all generalizations, no matter what their origin, must be repeatedly tested before being assigned a truth value, is another matter altogether: here we are dealing with the epistemological problem of the origin or mechanism of generalization, not with the methodological problem of validation.

Consider a classical or Pavlovian conditioning experiment. An animal is presented a light (or a tone) and a short time afterwards is given food (or an electric shock). After a few times the animal has learned to associate the two stimuli, so that on being subjected to the first (unconditioned) stimulus he makes the appropriate response (salivation, withdrawal, or what have you) before the second stimulus is delivered. The light (or sound) need not be always the same, and the food (or painful stimulus) can vary as well: the animal learns quickly to pair a whole class *S* of stimuli with an entire class *R* of responses. In mathematical terms, the animal has unwittingly built an entire *function* from *S* (actually from *S* cross the set of its internal states) to *R* out of a few values. (This function is "embodied" in a new configuration of neurons: recall Ch. 2, Section 1.1.) Consequently the animal has acquired a potential or propensity for reacting to further stimuli of the same kind with further responses of the same kind. In short, the animal has generalized—perhaps from a single case. This hypothesis explains instant learning, in particular instant generalization (in contrast to inductive or step by step extension). See Figure 5.6. Incidentally, this hypothesis takes the epistemological aspects of the problem of induction away from philosophy into experimental psychology and neuropsychology. The problem has now become that of identifying the neuronal systems and processes that do the generalizing, either prompted by a handful of cases or prior to observation.

We all know the advantage of generalization: it compresses a large (sometimes infinite) number of epistemic items. (And in the case of the transition of "some" to "all" we also gain in precision.) On the other hand by generalizing we may incur falsity. The whole of magical thinking

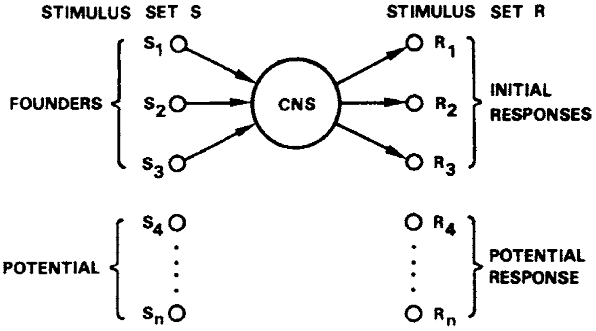


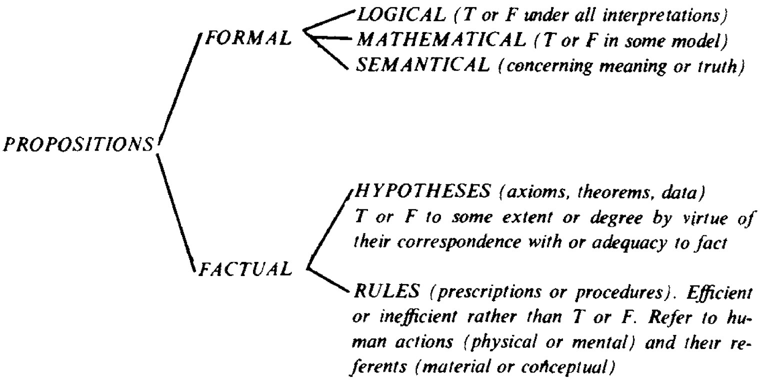
Fig. 5.6. A handful of stimuli are often sufficient to elicit the formation of a new psychon capable of producing new responses of the same kind beyond the initial ones.

derives from hasty (untested) generalization from coincidences. You came just when I was wishing you to come; I was not injured in the accident because I was carrying my amulet; my wrongdoing was not punished because I had paid my tribute to the deity. When such coincidences occur, our nervous system records them; when they do not, it does not. What is wrong about jumping to general "conclusions" is not the jumping itself—a beautiful instance of creativity—but the refusal to subject the generalization to tests before believing it.

2.2. *A Priori and a Posteriori*

What we take to be the basic distinction among propositions was drawn by Hobbes and elaborated on by Leibniz: it is the one between truths of reason and truths of fact. We generalize to half truths (and falsities) and speak in general of *formal* and *factual propositions*. The former are self-sufficient: they do not describe anything real; the latter describe (truly or falsely) some (possible or actual) fact(s). Thus a theorem in pure mathematics is a formal proposition. On the other hand a theorem in some theoretical factual science, like physics or psychology, though proved with mathematical rigor, rests on factual premises because it concerns real things. Besides, its truth or adequacy to fact must be shown by its consistency with the relevant empirical data.

There are several kinds of formal propositions and likewise several types of factual ones. We class them as follows:


PROPOSITIONS
FORMAL
LOGICAL (T or F under all interpretations)
MATHEMATICAL (T or F in some model)
SEMANTICAL (concerning meaning or truth)
FACTUAL
HYPOTHESES (axioms, theorems, data) T or F to some extent or degree by virtue of their correspondence with or adequacy to fact
RULES (prescriptions or procedures). Efficient or inefficient rather than T or F. Refer to human actions (physical or mental) and their referents (material or conceptual)

The preceding classification is more correct than either of the popular dichotomies analytic/synthetic or a priori/a posteriori. For one thing the concepts of analyticity and aprioriness have not been well defined except in extreme cases. Thus, a tautology is clearly analytic, but what about a theorem in theoretical physics, which has been derived by purely conceptual means? And an empirical datum is clearly a posteriori, but what about a factual hypothesis not built inductively and which happens to anticipate experience correctly? Neither of the two distinctions amounts to a dichotomy, hence neither is a sound principle of classification. When taken together the above classifications produce not only analytic a prioris and synthetic a posterioris but also analytic a posterioris and synthetic a prioris. Our own classification allows us to dispense with all four categories, in particular with the latter.

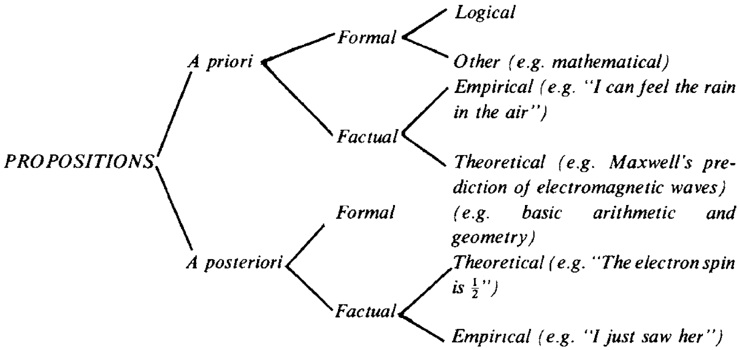
To be sure, it does make sense to use the expressions 'a priori' and 'a posteriori', but with the understanding that they give rise to subjective and shifting distinctions. Subject A conceives a given hypothesis on the strength of his experience; once conceived, the hypothesis functions as an a priori in A's further exploration of the world. Subject B conjectures the very same hypothesis (e.g. by analogy) and uses it to gain new experience (e.g. by employing it in the design of an experiment). Thus the hypothesis is a posteriori for A and a priori for B at a given moment in each subject's cognitive development; and the next moment what was a priori becomes a posteriori and conversely. In short, the definition "x is a priori iff x is prior to experience and independent of it" leads to contradiction. We must relativize it to the knowing subject and his experience, namely thus:

(i) Construct (concept or proposition) x is *a priori* for subject y at time $z =_{df} y$ has formed x by time z, and y has had no experience relevant to x before z (i.e. y knows x without resorting to perception). Otherwise x is *a posteriori* for y at z.

(ii) Construct x is *absolutely a priori* $=_{df}$ x is a priori for any subject at some time. Otherwise x is *absolutely a posteriori.*

These are *epistemological* concepts of aprioriness not to be confused with the *semantic* one of lack of real reference or the *methodological* one of validity regardless of experience. Whereas some mathematical and scientific constructs are epistemologically a priori, only the formal ones are methodologically a priori, i.e. they stand or fall regardless of any empirical operations. Absolutely a priori constructs, such as those of logical consistency, infinity, and topological space, are not to be confused with innate ideas of the kind postulated by Socrates, Freud, or Chomsky. In particular the constructs of logic and mathematics, though seldom originating in experience, cannot be formed without considerable prior experience of some kind. Only our nervous system, possessing certain dispositions built into its very organization, is innate. Thus it is possible that the tendencies to locate external events in space and time, and to search for causal connections, are innate, though perhaps not inborn. However, such dispositions do not constitute knowledge: they are only necessary conditions for learning.

The a priori/a posteriori dichotomy can be elaborated on. In particular, we must admit *factual a prioris*, such as the novel predictions made with the help of some scientific theories. (Think of the "missing links", now found and called *hominids*, or of the forecasts of World War II.) The resulting classification is this:


PROPOSITIONS
A priori
Formal
Logical
Other (e.g. mathematical)
Factual
Empirical (e.g. "I can feel the rain in the air")
Theoretical (e.g. Maxwell's prediction of electromagnetic waves)
A posteriori
Formal
(e.g. basic arithmetic and geometry)
Factual
Theoretical (e.g. "The electron spin is ½")
Empirical (e.g. "I just saw her")

3. EXACTIFICATION

3.1. *From Vagueness to Exactness*

Many, perhaps most of the concepts designated by ordinary language expressions are vague, i.e. imprecise or inexact: think of "small" and "most". Even in mathematics and science newly minted concepts are likely to be somewhat vague, and none becomes completely exact unless incorporated into a theory. Whereas in some cases vagueness is harmless and even expedites thinking—or at least speaking—in others it is a serious stumbling block to thinking and communicating.

Vagueness precludes the use of logic, which deals only with exact predicates. (In particular vague predicates violate the excluded middle principle, which rejects borderline cases by stating that nothing has and fails to have any given attribute.) Take for example the concept "most", the negation of which is "none or some or all". Now, ‘none or some" is tautological, for it amounts to "not some or some". Hence the negation of a most-statement amounts to the disjunction of a tautology and a universal statement. So, no matter what the truth value of the latter may be, the disjunction is logically true—which renders some-statements immune to tests. This shows that the ‘most’ of ordinary language violates logic. But the concept can be exactified, e.g. by construing "Most *A*'s are *B*'s" as "The relative frequency of *A*'s that are also *B*'s exceeds $\frac{1}{2}$". This construal avoids the above paradox.

Vagueness can often be reduced or even eliminated by some conceptual means or other, such as definition, quantitation, or theorization. Thus the expression ‘*B* depends on *A*’ can be rendered exact by construing *A* and *B* as sets, or members of sets, and dependence as a mathematical function from *A* to *B*. (For example, the ontological thesis that the past determines the present may be exactified as "For every thing, and every property of it relative to any given reference frame, there is a function representing that property, with respect to that frame, and such that the current value of the function is a function or a functional of its past values".) The vague concepts of fatness and baldness may be replaced with a certain standard weight value and a certain standard number of hairs respectively. The indefinite "some" may be eliminated either by quantitating or by dividing the original set into homogeneous subsets every member of which possesses the property designated by the predicate in question. And the ever present "etc." may be replaced with either an exhaustive list—which is not always possible—or a mention of the proximate genus. Thus instead of saying ‘spinach, broccoli, string beans, etc.’, we can say ‘green vegetables’, thereby gaining in both precision and breadth. Being a conceptual defect, vagueness can be reduced only by conceptual means: purely linguistic tricks will not help.

Vagueness is imprecision in meaning. Now, meaning can be construed as composed of sense and reference or, in self-explanatory symbols, $\mathcal{M}(c) = \langle \mathcal{S}(c), \mathcal{R}(c) \rangle$ (Vol. 2, Ch. 7). The sense of a construct in a given context equals the set of its logical relatives, i.e. implicants and implicates. And its reference class is of course the set of all the objects referred to by the construct. (Caution: Reference must not be mistaken for extension. A predicate referring to mythical objects refers and yet has an empty extension.) Since the meaning of a concept is composed of its sense and its reference, its vagueness can be of either. We shall therefore speak of intensional and referential vagueness. Moreover an imprecise sense may result in a somewhat indefinite reference class: think of the concepts of thing, organism, thought, and democracy.

The concept of *intensional vagueness* can be analyzed in terms of the full but unknown sense $\mathscr{S}(c)$ of a construct c, and the subset $\mathscr{S}_k(c)$ of known members of its sense, namely thus: $\mathscr{V}_i(c) = \mathscr{S}(c) - \mathscr{S}_k(c)$. But, since only $\mathscr{S}_k(c)$ is well determined, the concept of intensional vagueness is itself vague and therefore of little use. See Figure 5.7. On the other hand the concept of referential vagueness is quite clear-cut. A reliable test for the referential definiteness of a predicate is its performance in dichotomy: if a predicate allows us to unambiguously and exhaustively perform a partition of any set into the subset of all the members with the given property and the complementary subset of all those elements that lack it, then the reference class of the predicate is entirely definite. If, on the other hand, there are borderline cases that might as well be grouped with one category or its complement, then the predicate is referentially vague.

(A precise measure of such vagueness is provided by the fraction of borderline cases. In obvious symbols, $\mathscr{V}_r(c) = n/N$, a formula that works only for finite universes of discourse. It follows that $\mathscr{V}_r(\neg c) = \mathscr{V}_r(c)$, i.e. referential vagueness is shared to the same extent by a concept and its dual. And the *degree of definiteness* of c may be defined as the complement of that number to unity. Warning: These are just the degrees of referential vagueness or definiteness *known* at a given moment to a given knowing subject. Different workers may and, in fact, usually do, come up with different counts of borderline cases—a common occurrence in ordinary life as well as in biological and sociological systematics. So, although the concept of referential vagueness is quite precise, its application varies from

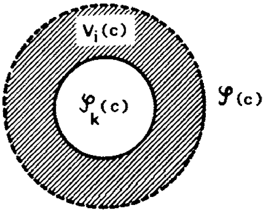


Fig. 5.7. Intensional vagueness: the ill-limited shaded zone.

subject to subject. Finally, the *total vagueness* or *indefiniteness* of a concept c may be defined as $\mathscr{V}(c) = \langle \mathscr{V}_i(c), \mathscr{V}_r(c) \rangle$.)

Intensional vagueness is a necessary but insufficient condition for referential vagueness. I.e., if a concept is referentially vague then it is also intensionally vague but the converse need not be true. In fact intensional vagueness can often be palliated by introducing and using practical *criteria* (decision rules) enabling one to decide whether a given object belongs to the reference class of the concept in question. Thus long before any of the essential properties of gold were known, chemists used *aqua reqia* for identifying golden things. And, although nobody knows what scientific talent is, anybody can spot it by counting numbers of tough solved problems, or even publications.

However, in scientific contexts we cannot be satisfied with referential precision: we also need intensional definiteness. This is best achieved by turning the given context into a member of a mathematical theory. However, mathematization has its limits: (*a*) it is only necessary, not sufficient, as shown by the case of fuzzy set theory, and (*b*) it concerns only the form or structure of a concept, not its factual reference. In any event, regardless of the means whereby a construct is sharpened, we may stipulate that a construct is *exact* if and only if it is not vague—i.e. if both its intensional and its referential vagueness are nil.

To *exactify* a concept, then, is to replace it with another having a precise sense and a precise reference, each of which has a sizable overlap with the sense and the reference, respectively, of the original imprecise concept. A perfect coincidence of sense and reference cannot be expected: every time a concept is exactified something is gained and something else—usually simplicity and intuitiveness—is lost. Think of the contemporary concepts of number, matter, and life, as compared with their predecessors.

Needless to say, exactification is not identical with quantitation: there are numerous qualitative mathematical disciplines, such as logic, abstract algebra, and topology. Nor is exactification restricted to the natural sciences: there is a vigorous exactification movement in the social sciences and the humanities, not excluding philosophy. (See, e.g. Bunge (Ed.), 1973d.)

Concept refinement, or exactification, is one aspect of the progress of knowledge. In fact the latter does not consist in piling up new data and new conjectures, but in a selective process that pivots around the invention and improvement of concepts, propositions, theories, and methods. In this process concept refinement plays an indispensable part: although it does not replace the original invention of an initially imprecise idea, it helps to develop it. This must be insisted on because the role of concept refinement has been exaggerated by some and underrated by others. The overrating has been mainly the work of theorists, which is understandable, for they are professional concept breeders. And the underrating is not surprising either in view of the empiricist tenet that genuine research is concerned only with fact finding. This one-sided view was justified at the beginning of the modern era, as a reaction against the hollow verbosity of decadent scholasticism. Recall that Bacon advertised his own philosophy as "a philosophy of works, not of words", and that the Royal Society of London took from him its motto *Nullius in verba*—"There is nothing in words".

The anti-intellectualist tendency became obsolete no sooner it was born, for modern science and the technology based on it are as conceptual as they are empirical. In fact many problems in science and modern technology are problems of concept formation or refinement, which occur in the building, application, and evaluation of theories. Think of the two centuries of scientific debates over the term 'force', which turned out to designate now the concept of force proper, now that of energy, now that of power. Or, to turn to present day problems, think of the notions of mind, creativity, and intelligence, which still roam wildly in psychology. Or of the concepts of economic value, money, and social class, paragons of inexactness in social science. The vagueness of these notions hampers theoretical development and empirical research: it only encourages confusion and barren argumentation. (See an unabashed praise of imprecision in Feyerabend 1981, p. ix.)

What can be done about inexact concepts? The tradition in mathematics and natural science has been to replace them with exact concepts. In recent times a new strategy has been advocated, namely that of accepting vague predicates and treat their ill defined extensions as fuzzy sets. (An exact predicate has an exact extension, i.e. one the membership of which is definite. An inexact predicate has a fuzzy extension, i.e. one the membership of which is not well defined; hence, given an individual, one may not be certain whether or not it belongs to the set. The mathematical theory of fuzzy sets has been invented to cope with such collections.) The decision to use fuzzy sets or abstain from using them in representing reality is a methodological one. The usual arguments are that reality itself is fuzzy, and that fuzzy concepts are respectable after all since we can speak about them in an exact way with the help of the theory of fuzzy sets. Let us deal briefly with these two arguments.

There is no evidence that reality itself is fuzzy or indeterminate. What is true is that certain exact concepts do not apply to some things: for example, the concepts of position, trajectory, and shape do not refer correctly to elementary particles. In such cases one introduces different but still exact concepts, such as that of position distribution instead of sharp position. Moreover all the successful factual theories have formalisms presupposing the theory of ordinary sets. In short, it is not that the world is fuzzy, but that our knowledge of it is imperfect (but perfectible). As for the claim that fuzzy sets have become respectable for being the subject of a mathematical theory, it is like saying that crime became honorable since the inception of criminology. Using fuzzy sets in modeling reality is a confession of defeat. The point is not to be able to speak exactly about inexactness but to reduce the latter and, if possible, to eliminate it altogether. Dirt must be recycled or eliminated, not be shoved under the carpet, let alone exhibited proudly. "Everything that can be thought at all can be thought clearly. Everything that can be said can be said clearly" (Wittgenstein, 1922, 4.116).

3.2. *Quality and Quantity*

Primitive and archaic thought are characteristically qualitative and analogical (Lloyd, 1966; Goody, 1977). By employing exclusively qualitative or dichotomous predicates, primitives and ancients were necessarily led to dichotomous or polar classings: cold and hot, heavy and light, clear and dark, near and far, wild and domesticated, good and bad. This mode of conception misses nuances and is apt to lead to a Manichean or a dialectical world view. Scientific and technological thinking, on the other hand, presumes that most properties come in degrees—of heaviness and speed, heat and nearness, excitation and value, and so on. Only in some backward branches of the social sciences and the humanities most thinking remains qualitative, hence polar or black and white.

True, a few properties are objectively qualitative, hence correctly representable by dichotomous predicates. Existence is an, perhaps *the* example: a thing either exists or it does not—there are no degrees of existence. (For the construal of existence as a predicate see Vol. 3, Ch. 3.) But most, perhaps all other properties, come in degrees that can be ordered or even assigned numbers. For example, there are different degrees of stability, including metastability and ultrastability. And, of course, there are infinitely many degrees of size, speed, probability, and what have you.

We cannot know a priori whether a given property is inherently qualitative or will turn out to be quantitative. We can only try to form a quantitative conceptualization of the given property and measure the latter to see whether we get more values than one (presence) and zero (absence). In any case modern science and technology since Galileo have adopted the methodological postulate that *every property, except for existence, can be quantitated.* Faith in this postulate has sustained all the investigators who have succeeded in quantitating properties previously believed to be imponderable. The postulate is confirmable but irrefutable for, should a given property resist our efforts to quantitate it, we may blame ourselves for our failure.

Having praised quantitation we must warn against the temptation to grade even complex properties by single numbers. Such oversimplification results in adopting IQ as an overall measure of intelligence, and increase in GNP as an overall index of progress. When such simple-minded quantitations are shown to be inadequate, irrationalists are confirmed in their mistrust of numbers. Better not give them this pleasure and recognize that in many cases a scalar will not do, but that a vector, a tensor, or a matrix may be needed.

Evidently quantitative knowledge, when available, implies qualitative knowledge but not conversely. Thus "distance" subsumes "separateness", "rate of change" subsumes "changing", and "probability" subsumes "possibility". If a quantitative pattern (law) has been found, its qualitative features can be extracted. Thus whether the evolution of a system is stable or unstable, can be found out theoretically only after its equation of evolution has been set up and tested—which equation involves numerical functions. (This is how Poincaré proved the stability of the solar system.) Therefore it is not true that the study of qualities always precedes that of quantities. The deepest qualitative knowledge is a byproduct of some quantitative knowledge. All of the social sciences could profit by this maxim; some of them, in particular economics and socioeconomic history, are already applying it.

The fundamentals of quantitation are these. Let S be a set of degrees or intensities of some property—objective degrees such as temperatures or subjective ones such as hotness. Postulate that S is simply ordered by a relation $\gtrsim$. (Hence if x and y are in S, then $x \gtrsim y$ or $y \gtrsim x$, and both statements hold only in case $x \sim y$.) In most cases this comparative concept is insufficient. We *quantitate* the given concept by mapping S onto a set of numbers in such a way that (*a*) each degree (member of S) be assigned a single number (one to one correspondence), and (*b*) the order in S be preserved, i.e. the numerical image of S reproduces the objective order in S.

(In other words the quantitation of S consists in introducing a mapping $M: S \rightarrow T$, where T is included in the real line, such that, for any x and y in S, $x \gtrsim y$ if and only if $M(x) \geq M(y)$. Thus "farther away than" will be quantitated as "at a greater distance than", and "hotter than" as "having a greater heat content than". In general such mapping is not unique: i.e. in principle there can be different quantitations of one and the same property. Think of the different temperature scales and the different modes of reckoning volumes of production. The choice among the various possibilities is dictated in some cases (e.g. length) by convenience, in others (e.g. temperature) by theory. If S is or is assumed to be nondenumerable, so must be its image T even if we do not possess the technical means for measuring the difference between two close members of T. If the map M is one to one, order preserving, continuous, and has a continuous inverse, it is called a *homeomorphism*.)

(The above procedure presupposes that we are given the entire set S of degrees of some property. This is sometimes the case. As a rule we build a quantitative concept regardless of empirical findings, and design observations or measurements in order to find some of the values of the function concerned. More precisely, we introduce a function $M: D \rightarrow T$, where D is in general a cartesian product of certain sets. At least one of these is a set of concrete things, such as bodies or societies, and at most one other factor is a set of units, such as the multiples and submultiples of the meter. For example, if we wish to reckon or measure distances between things in the context of relativistic physics, we introduce the distance function $d: P \times P \times F \times U_d \rightarrow \mathbb{R}^+$ obeying the usual distance axioms, where P is a set of physical objects, F that of reference frames, U_d the set of distance units, and $\mathbb{R}^+$ the set of positive real numbers. The mathematical theory of metric spaces leaves P uninterpreted and fails to include F and U_d. Any map $M: D \rightarrow T$ with the above characteristics is called a *quantitative predicate* or *magnitude*. See Bunge (1971, 1973a).)

(Qualitative predicates are particular cases of quantitative ones. Indeed the former are functions of the form $Q: D \rightarrow \{0, 1\}$ such that $Q(x) = 1$ iff x possesses the given property, and $Q(x) = 0$ otherwise. Once we have a set of magnitudes, e.g. those occurring in a mathematical model of a real system, we may focus our attention on certain global or qualitative features of the system. To this end we may need to *discretize* those magnitudes, i.e. to replace them with functions ranging over the integers. A common discretization is that of time, particularly when computers are employed. In this case differential equations are replaced with finite difference equations.

The discretization of a continuous system model is not always a trivial exercise. Sometimes it is suitable to discretize the state variables themselves, as when one stipulates that any values below threshold count as 0, and any above threshold count as 1. At other times one is interested in growth and decline rather than in absolute values. In such cases the rates are discretized, namely thus: the state of the variable is assigned the number 1 if its rate is positive, and 0 if negative. In this case the momentary state of the system is represented by an *n*-tuple of zeros and ones exhibiting the state of each variable: Glass (1975). In short, once again the qualitative features of the system are inferred from a quantitative model of it.)

Every real property is representable by at least one mathematical function, but the converse is false. Genuine quantitation obeys a methodological condition as well as a mathematical one. It must be accompanied by an indication that there are (direct or indirect, actual or potential) ways of effectively assigning (measuring) some values of the function in question. Otherwise quantitation will be judged phoney. Three examples should suffice. One is Herbart's phoney mathematical psychology. Another is the widespread belief, among contemporary philosophers, that propositions can be assigned probabilities, even though they are very discreet with regard to the method whereby such assignment is to be made. A third example of phoney quantitation is this. One can easily speculate that dreams have densities, and can imagine some postulates defining the dream density function—e.g. if dream *a* precedes dream *b*, then the density $d(a)$ is greater than the density $d(b)$. But this number game has no substance because we do not have the faintest idea how to determine effectively $d(a)$ and $d(b)$.

So much for quantitation, which we shall meet again when studying measurement (Ch. 11, Section 3.1). The two operations should not be confused, although they often are in the methodological literature (e.g. in Suppes and Zinnes, 1963). Quantitation is a purely conceptual operation, whereas measurement is empirical as well as conceptual. We can have quantitation without measurement but not conversely. For example, the concept of power exerted by organized crime in a given society may be quantitated, e.g. as the number of people whose economic or political activities are controlled by it, or by the annual volume of its business. However, since organized crime is supposed to be illegal, no reliable statistics can be compiled to measure its power. Still, this power is real, it can be roughly estimated, and one day it may be discovered. On the other hand it is impossible to measure in any objective way the probability of propositions, in particular hypotheses, so assigning them probabilities is a prime example of bogus quantitation. Although quantitation is different from measurement and precedes the latter, it is pointless unless measurement is possible at least in principle.

4. REPRESENTATION

4.1. *Conceptual Map*

Percepts tend to cluster into systems that map the body or the environment: those systems are the perceptual maps, which allow the animal to move around, make a living, and defend itself (Ch. 4, Section 3.1). Concepts also tend to organize into systems: some of them map the body, others the environment, and finally some (namely the theories in formal science) map nothing at all. All animals endowed with a central nervous system form perceptual maps, but only the higher vertebrates seem capable of forming conceptual maps, i.e. representations that overreach perception. Unlike the perceptual maps, the conceptual ones allow the animal to understand, forecast, and plan to some extent.

All conceptual maps are patterns of neural activity. However, some are externalized and thus become perceptible to other subjects: this is the case with diagrams, geographical maps, and systems of sentences or utterances. Whether or not externalized and thus visualizable or audible, conceptual maps summarize and systematize conceptual knowledge (or prejudice). And, whether or not they are accurate (true) representations of real items, they help us orient ourselves (successfully or not), as well as to act in the real world, in at least as effective (or misleading) a manner as perceptual maps.

Consider diagrams for a moment, and in particular the diagrammatic representations of a concrete system with known of conjectured composition, environment, and structure. About the simplest such representation is a diagram exhibiting the components (at a given level), the (relevant) environmental items, and the (main) relations among all such items: see Figure 5.8. Note that, although the diagram is perceptible, it is a conceptual or symbolic map not a perceptual one, if only because it may represent imperceptible things, such as an atomic nucleus composed of two nucleons held together by imperceptible links, such as fields. Note also that such a diagram is an extremely poor representation of any system: it summarizes only the minimal body of knowledge necessary to speak about a system. A stack of diagrams of the same type may provide a more detailed

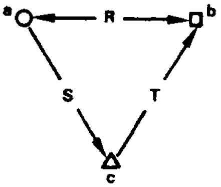


Fig. 5.8. Schematic representation of a system with two components, *a* and *b*, held together by the link *R*, and interacting with environmental item *c* via couplings *S* and *T*.

representation. But only a theoretical model of it—e.g. a system of differential equations—can give a detailed and deep representation. (We shall deal with theories in Ch. 9.)

Here are a few principles concerning the diagrammatic representation of material things:

(i) A diagram of a system should represent its composition, environment, and structure, either directly or in terms of properties. (A mixed diagram, containing arrows from things to properties or conversely, makes no sense because one and the same symbol, namely the arrow, is now made to denote a link or coupling, now the subject-predicate relation. However, such diagrams occur frequently in the literature.)

(ii) One and the same concrete system may be represented in alternative ways, and any such alternatives may or may not be equivalent. (They are so only if they can be obtained from one another via a topological transformation, i.e. one preserving the relations of vicinity.)

(iii) If two diagrams representing a given system are inequivalent, then they complement each other or one of them is more accurate than the other.

(iv) Diagrams are no substitute for theories. In particular, although they may suggest explanations, they cannot produce forecasts.

Like perceptual maps, conceptual maps can be analyzed as partial functions from stimulus events to neural events. (Cf. Ch. 4, Section 3.1.) But, whereas perceptual maps are limited to appearances, conceptual maps may represent reality, i.e. they may be subject-invariant. Think of a coin, its possible appearances to a subject, and the conceptual models of it. A single coin may appear to us in any one of infinitely many fashions—as many as its relative positions to ourselves, dynamic states, illumination conditions, etc. On the other hand only a finite number of conceptual models (e.g. geometrical, dynamical, and probabilistic) of a real coin can be built, none of which involves a subject.

Another difference between the two kinds of map is that, whereas perceptual maps resemble the represented objects, conceptual maps need not hold such a relation of analogy with their objects or referents. Think of a family tree, or of a system of econometric equations: unlike blueprints and geographical representations, such symbolic maps do not resemble their referents. Therefore the words 'picture' and 'reflection' do not describe them adequately. This suffices to write off naive realism (the copy or picture or mirror view of knowledge) as inconsistent with science and technology.

Contrary to naive realism, Helmholtz held that all our representations are symbolic and that even perceiving involves conjecturing. He wrote that representations are not copies or images but "only *signs* [symptoms] of external objects, and in no sense images of any degree of resemblance. [...] For a *sign* it is sufficient that it become apparent as often as the occurrence to be depicted makes its appearance, the conformity between them being restricted to their presenting themselves simultaneously; and the correspondence existing between our sensations and the objects producing them is precisely of this kind. They are signs which we have *learned to decipher*, and a language given us with our organisation by which external objects discourse to us—a language, however, like our mother tongue, that we can learn only by practice and experience" (1873, pp. 391–2).

Conceptual maps, and particularly factual theories, are indeed symbolic in some respect or other and so need not resemble their referents: Helmholtz's view is true of them. On the other hand it does not hold for perceptual maps: if the latter did not preserve at least roughly (some of) the relations among the stimulus events, we would be unable to move around and act in our environment. So, we accept Helmholtz's view for conceptual maps but not for perceptual ones. Incidentally, that view, though at variance with naive realism, is consistent with critical realism and even with materialism, for it presupposes the autonomous existence of the external world: Helmholtz insists that our ideas are signs of the latter. (This point was misunderstood by Lenin (1908, pp. 237ff.), who thought that if our ideas are only signs then the existence of external objects becomes doubtful.)

Naive realism is allied to mechanism, according to which the world is a mechanical system, so that every real thing can be pictured literally, i.e. described "*par figures et mouvements*" (Descartes). Since the inception of field physics we have learned that this view is overly optimistic; quantum mechanics has weakened further the mechanistic thesis, and the social sciences have never had any use for it. We know that only some physical objects, in particular medium sized mechanical systems, can be represented iconically, and even so only in a few respects. Electrons and photons, biopopulations and societies, cannot be diagrammed in detail: only symbolic systems can represent them. This does not mean that they are less real than stones, but that they have properties that cannot be grasped by perception. But, of course, the theories representing such things can be externalized by means of written symbols, and thus become vicariously perceptible.

Once a conceptual map has been drawn, written down, or taped, it becomes a cultural artifact, i.e. one that can be examined, modified or used by others. By becoming public, the map seems to acquire an existence of its own. But this is an illusion, for the drawing, printed page, or tape, would not perform any functions in the absence of brains. In other words, we have to do with three objects, and therefore three relations: the represented thing, its conceptual map in some brain, and the public manifestation of the latter. The relation between the latter and the thing it maps is a sort of product or composition of the relations artifact–brain, and brain–represented thing. See Figure 5.9.

4.2. *Discovery or Invention?*

If conceptual maps are not mere copies but instead symbolic constructions, then they must be creations like poems or musical scores and unlike photographs or tapes. Hence when exploring our environment we do not always make discoveries, as when finding a coin in a pocket, but sometimes we invent beforehand or at the same time that we discover, i.e. we discover a thing and invent a conceptual map of it. Thus Columbus was able to discover new lands for the Europeans only because he had previously

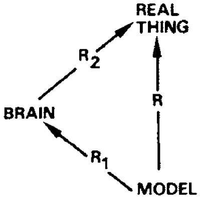


Fig. 5.9. The modeling relation R equals the composition of the model-brain and the brain-represented object relations.

formed a (rather inaccurate) conceptual map of the planet. When Harvey announced that he had discovered the circulation of the blood, he built a conceptual map of the human cardiovascular system complete with capillaries, which he had been unable to see. (So, the inimitable 17th century gossip John Aubrey (1949) was not totally wrong when he described Harvey as "Inventor of the Circulation of the Bloud"—as a result of which "he fell mightily in his Practize, and 'twas beleeved by the vulgar that he was crack-brained.") And when J. J. Thompson discovered the electron in 1898, he produced a conceptual map of it as a tiny electrically charged corpuscle. Others might have thought of cathode rays as jets of an electrically charged fluid: they would have missed discovering the electron.

Every important discovery is accompanied by the invention of a conceptual map of the object discovered. Moreover such a map is what makes the discovery important. Thus if an ignorant person finds a primitive tool and does not recognize it as such, he makes one more discovery among the many he makes every day. If the same object is found by a competent archaeologist, he may fit it into a conceptual map of a primitive society, and thus turn it into an important discovery. Likewise when an inventor imagines new uses of known things, assembles old units into new systems, or designs totally new systems, he forms conceptual maps. And some of his inventions may help make new discoveries. What holds for the invention of artifacts holds also for mathematical invention. Even the discovery of a previously unknown consequence of a set of premises, as in the proof of a theorem, is a creative process, for it leads to something new. A fortiori, the invention of new concepts, new theories, or new theorems is a creative process, and moreover an original one.

People are creative all the time: they keep on discovering or inventing things or ideas, mostly ordinary, sometimes extraordinarily clever or silly. But not all creations are original: many are rediscoveries or reinventions. A creation is said to be *original* when nobody but its creator knew it before. Obviously, original creations are not equally valuable: whereas some are good, others are bad and most are indifferent. Hence originality is not good in itself: only good originality is. Nevertheless some societies encourage originality in itself whereas others discourage originality and thus block progress—which is understandable, because novelty may threaten any rigid social order.

The notion of creativity is suspect in many quarters, in particular in empiricist philosophy and its psychological counterpart, behaviorist psychology. According to these doctrines all the organism can do is to admit environmental inputs, in particular pick up the information inherent in the environment. This environmentalist thesis would never have occurred to a historian: he knows that theories and machines are invented, not gathered like wild fruit or fish. Nor does the neuropsychologist have any difficulty in admitting creativity: to him it is the ability to form new neuronal configurations (Bunge, 1980a). This ability is partly innate, partly fashioned (or deformed) by education. A dogmatic education will systematically punish originality and encourage blind repetition, whereas a progressive education is supposed to free creativity—though it sometimes fails to promote it for fear of falling into dogmatism.

In general, all the higher vertebrates are creative, and a few individuals are original in some way or other. There is nothing mysterious about originality although there is much to learn about it. Every one of us has a peculiar brain composition and organization, and every one of us is subjected to environmental circumstances that differ somewhat from those of other individuals. (In particular, no two individuals have exactly the same protein composition and consequently can engage in exactly the same neurochemical reactions. And no two humans are treated exactly alike by their fellow humans or are presented with exactly the same opportunities and challenges.) In conclusion, there is a grain of truth in empiricism: we cannot escape our environment although we can modify it somewhat. And there is a grain of truth in rationalism as well: we create our own ideas, in particular conceptual maps—though not out of the blue. So, the whole truth contains a grain of empiricism and another of rationalism. However, it is not a mere merger or synthesis of the two, for each of them ignores the brain and that very particular environment which is society. We shall return to this subject in Ch. 15.

We value creativity but do not know the secret of it. Nobel laureate Linus Pauling was once asked how he had managed to propose so many good original ideas in so many different fields. He replied: "It's simple. You just have plenty of ideas, and cast off the bad ones". Yes, but how do you come by plenty of ideas, good or bad? And how do you manage to spot the bad ones? The first question is yet to be answered by the psychology of thinking; the second will concern us in Ch. 12.

5. CONCLUDING REMARKS

Perception gives us only perceptual knowledge, which is egocentric and limited to appearances. Only conceptual knowledge can be objective and deep: only conceptual maps give us a glimpse of things in themselves. (Cf. Schlick, 1925.) Still, conceptual knowledge is not independent of perception, which sets conception in motion and checks its products. This view, which can be found *in nuce* in Aristotle, is part of every critical realist epistemology.

Critical realism takes some elements from empiricism and others from rationalism but goes beyond both of them. Empiricism has always extolled the virtues of perception and rejected all concepts that did not originate in it. On the other hand rationalism has always praised the superiority of conception to the point of proclaiming the self-sufficiency of reason. Critical realism is a sort of synthesis of empiricism and rationalism, but it is not the only possible synthesis. Recall that Kant had managed to put together the negative aspects of empiricism and rationalism, by holding that we can have no experience without certain a priori intuitions, that things conform to human thought rather than the other way round, and yet we can know only phenomena, not things in themselves—all of which is out of tune with the realistic epistemology inherent in modern science and technology. (More in Ch. 15, Section 3.)

Although we pride ourselves on being the only animals capable of certain conceptual feats, such as building mathematical theories of the world, we still know very little about the psychology and neurophysiology of conceptual operations. This backwardness seems to have two main causes. One is our familiarity with such processes. (Familiarity may not breed contempt, but it certainly fails to spur inquiry.) The other is the belief that those operations could not possibly be studied scientifically—a belief nurtured by behaviorism. However, at last the scientific study of conception is under way since about 1960. Besides, we have learned to analyze many conceptual processes and have set up a number of rules for the correct performance of some of them, such as classification and deduction. Paradoxically, methodology is somewhat more advanced than descriptive and explanatory epistemology: we know how to perform certain operations correctly while having only very rough ideas of how we manage to do it.

Conceiving is anything but copying: it is a creative process going far beyond perception—which in the higher vertebrates is creative too, partly because it gets mixed with conception. We know very little about the mechanisms of conceptual creation, and this partly because of the empiricist tenet that every conception is a mere combination or elaboration of percepts. The classical examples are those of the centaur and the mermaid. The myriads of counterexamples supplied by logic, mathematics, science, technology, and the humanities, are merely ignored. Yet the invention of syllogisms, equations, grammars, the idea of an atom, and the wheel, owe nothing to perception. All we know is that conception is a brain process: we ignore its mechanism, and a fortiori the secret of originality.

CHAPTER 6

INFERRING

Conceptual knowledge can be augmented in either of two ways: by forming new propositions out of nonpropositional material (e.g. percepts), or by inferring new propositions from a set of known ones. We do the former, for instance, when producing empirical data and proposing conjectures to explain them. And we infer when, for example, drawing conclusions from a set of premises and leaping to generalizations on the strength of particulars. Both processes are equally important in all fields of inquiry. However, the importance of inference is often underrated by those workers who, misled by empiricism, believe that all they have to do is collect solid facts and abstain from committing what they regard as the capital sin of the intellect, namely jumping to conclusions. But, of course, those who do not jump do not overcome any hurdles, and so do not participate in any original cognitive venture.

As in the case of propositions, it is possible to adopt either a psychological or a philosophical (and in particular logical) viewpoint concerning inference. That is, we may regard it either as a brain process without paying attention to its validity, or as a timeless relation between propositions and focusing on its validity with total disregard for its neurophysiology. In line with our program of fusing the scientific and the philosophical approaches to inquiry, we shall adopt both viewpoints. However, we shall spend far less time on the former than on the latter because we know very little about the physiology and psychology of inference.

Every inference is either deductive or nondeductive. The former is demonstrative, the latter nondemonstrative and at most suggestive. Deduction is of course the central theme of logic, which has become a separate branch of mathematics. We shall have little to say about it. Nor shall we spend much time on nondemonstrative inference, in particular induction and analogy. Not that induction and analogy are unimportant, as the rationalists claim: they are pervasive, though perhaps not more so than intuition. Moreover they can and must be studied, though not with the hope of coming up with exact rules or methods, such as the rules contained in works on inductive logic and the much-vaunted inductive method.

Likewise love and hate deserve being studied, though not with the aim of producing scientific love or scientific hate.

The above does not entail that all nondemonstrative inferences are worth the same, namely little, let alone nothing. Nondeductive inferences, like guesses, can be wild or educated, gratuitous or compatible with data or theory. But whether or not they happen to be lucky is for tests, not for an a priori discipline, to decide.

1. FROM SPONTANEITY TO FORMALITY

1.1. *Natural Reasoning*

The psychobiological investigation of reasoning patterns is a recent development which is yielding few but important results on the thinking habits of children, primitives, and adults living in modern societies. (See Nisbett and Ross, 1980.) We shall make a quick review of some typical findings.

In all inference we must assume some premises, even if we do not know whether they are true, if only for the sake of argument, in particular to find out whether such assumptions lead to acceptable consequences. This detached attitude is not universal. The natural thing to do is either to accept or to reject a premise rather than to examine it with a purely cognitive concern. Entertaining or examining assumptions without committing oneself a priori to them is not done in early infancy and may be a rather recent historical acquisition. At least, the historical record does not seem to go farther back than a few centuries before our era.

In particular, children do not seem to be able to form conditionals, i.e. propositions of the form "If A then B", until about age 3, and this only when prompted by their teachers. Moreover when they do form a conditional they presuppose its antecedent A to be true: it is only much later that they can envisage conditionals with false antecedents or antecedents with unknown truth values. Likewise nonliterate peoples seem incapable of forming such conditionals. (Cf. Luria, 1976.) In short, hypothetical reasoning comes rather late or not at all: it seems to be peculiar to an advanced stage of culture.

Not surprisingly, argument too comes rather late. Piaget (1955) found that children go through two stages with regard to argument: quarrelling and argumentation proper. The former consists in a clash of views, and it is the mental equivalent of fighting. Genuine argument, which exhibits reasons, motives, or causes, does not appear until age 7. (But, presumably, the majority of such rational arguments are formally invalid.) Another result of Piaget's studies (1955) is that, before the ages of 7 or 8, children may reject self-contradictions but they do not often notice contradictions between opinions voiced successively: their body of beliefs does not form a system. Between 7 and 8 they start noticing and rejecting contradictions, thus building gradually belief systems or world views. (Whether this process is universal or peculiar to advanced cultures, is not known.)

Once a person has attained this rational stage she tries to eliminate contradictions within her body of beliefs by discarding some of them. However, if such contradictions persist—if the person cannot give up one of the theses in conflict—the individual tries to "rationalize" them. If she fails, she feels a psychological discomfort. And, as long as she feels such discomfort, she acts so as to reduce "cognitive dissonance", much in the same way that she tries to reduce other imbalances, such as hunger. One way of reducing dissonance is to avoid situations where the individual is forced to choose between the horns of her dilemmas. Social intercourse is at the same time a source of cognitive dissonance and a vehicle for eliminating it by introducing new knowledge. The greater the dissonance, the stronger the social pressure to reduce it. (Festinger, 1957).

Although the theory of cognitive dissonance does not seem to have been established experimentally (Chapanis and Chapanis, 1964), casual observation confirms that ordinarily we abhor contradictions but, when we cannot get rid of them, we atone for them in some way or other. For example, the contradictions inherent in every religion lead some people to claim that they are mysteries, others to patch them up with further absurdities. And the contradictions between dogma and action—e.g. preaching Christian love and waging war in the name of it—could always be solved, or at least absolved, by penance, fines, or service of the church's worldly interests (Tuchman, 1978).

Another recent finding of cognitive psychology concerns multiple reasoning in educated adults. It is the hypothesis that human thinking, far from being sequential, is multiple, in that it consists in a number of parallel threads grouped around a main sequence, which is the one within the conscious area: "Awake or asleep, a number of more or less independent trains of thought usually coexist. Ordinarily, however, there is a 'main sequence' in progress, dealing with some particular material in step-by-step fashion. The main sequence corresponds to the ordinary course of consciousness. It may or may not be directly influenced by the other processes going on simultaneously" (Neisser, 1963). Introspection confirms this conjecture.

There is a century-old debate on primitive mentality. The early anthropologists noted the great disparities between the savage mind and ours: the poverty of its conceptual stock, its difficulty in generalizing, its incapacity to entertain hypotheses, the clumsiness of its reasoning, and the general lack of interest in abstraction. Being often adjuncts to imperialist expeditions or colonial administrations, some of those anthropologists tended to attribute such backwardness to an inherent biological inferiority. The early liberal reaction to this racist claim was to deny that there is such a thing as the savage mind, and to hold instead that there is only less information: reason is universal, knowledge is not.

Later studies have shown that each of the two contenders held a grain of truth: there is primitive mentality but it has to do with schooling not race. The primitive tribesman cannot perform complicated conceptual operations because he has a poor conceptual kit and does not handle pencil and paper—but he can learn. In particular, knowing only a few numbers, he cannot keep track of any extensive stock, let alone perform calculations, e.g. of distances or speeds—but he can learn.

The anthropologist Sylvia Scribner (1977, p. 486) proposes the following generalizations on the basis of studies concerning proficiency in solving logic problems. "1. *In all cultures*, populations designated as 'traditional' or 'nonliterate' have just somewhat better than a chance solution rate across all types of problem material. [. . .] 2. *Within each culture* there is a large discrepancy of performance between schooled and nonschooled. The major jump seems to occur at levels of education as low as two to three years of school. [...] 3. With schooling, there is little *between-culture* variation in performance for the cultures studied. Grade, rather than society, is most determinative of performance". In short, although actual reason is not universal, potential reason is, and it can be actualized by schooling. So, there is a chasm between the savage mind and the civilized mind, but it is being crossed all the time with the spreading of schooling.

Even in civilized societies most adults seem to reason fallaciously most of the time as soon as they go beyond practical matters. Thus in a typical experiment only 9 percent of the subjects concluded correctly that the premises "All P are M" and "All S are M" do not entail anything about the relation between Ss and Ps (Chapman and Chapman, 1959). In another typical experiment, only half of the subjects drew the correct conclusion (namely "C is worse than A") from the premises that B is worse than A and C is worse than B (cf. Mayer, 1977). When faced with a set of data, the typical undergraduate conconcts a single hypothesis, is reluctant to consider falsifying data, and fails to consider alternative hypotheses (Johnson-Laird and Wason, 1977). Moreover, casual observation shows that resistance to give up a hypothesis in the face of adverse data mounts with the importance of the hypothesis in one's world view.

Typically, the layperson and occasionally also the scientist and the technologist operate with biased samples—usually collections of items that happen to be noticed and remembered—rather than with random samples; they detect, or fail to detect, covariations suggested by prior beliefs; they tend to expect and forecast what they wish to happen; they tend to pooh-pooh exceptions and criticisms; and generally they adopt beliefs on the strength of first impressions, authority, or fashion. (For these and other fallacies see Nisbett and Ross, 1980.)

The conclusion is obvious: logic is extremely artificial. However, it is not much more so than cooking, building, literacy, or reckoning. Right thinking does not come more naturally than reading or plane piloting: it must be learned. Cognitive psychologists have known this for half a century, and they have also known that most college students will consistently incur fallacies of all kinds unless they have taken courses in logic or in higher mathematics. However, knowledge does not go easily across disciplinary barriers, and so the vast majority of college students graduate without having been exposed to either logic or higher mathematics. And once habits of thought have been acquired it is hard to change them. Thus whereas the person who has been trained in mathematics tends to form clear ideas and reason correctly, the person who has been trained in the humanities is often tolerant of inexactness, and the individual who has had a theological or in general ideological training cannot help thinking in opposites and prefers criticism and debate over constructive reasoning.

1.2. *Formal Reasoning*

Spontaneous or natural reasoning is tacit, undisciplined, and often invalid. Formal reasoning is, by definition, explicit, regimented, and valid. It is formal because its validity is independent of the meaning of the premises and the conclusion. It is explicit in that it involves only explicitly stated premises and inference rules. It is regimented in the sense that it is subject to (though not directed by) principles or rules (e.g. those of the predicate calculus). And it is valid in that, when its premises happen to be true, its conclusions are true as well. (A rule of inference can be construed as a truth preserving function from n-tuples of propositions to propositions.) So far, only deductive inference has been formalized. There is no reason to hope that other types of inference (e.g. inductive and analogical) will follow suit, for they are content-dependent, i.e. their success depends on the nature of the case.

In ordinary language we often confuse reasons with causes: 'because' can be interpreted either as a causal relation (by cause of) or as a logical relation (by reason of). At first sight this is a confusion: reasons are conceptual objects whereas causes are events. (Bunge, 1959a.) However, on second thought there is something to be said for such conflation: ultimately every reason *is* a brain process and therefore a cause of some other bodily process. So, although not all causes are reasons, every reason is a cause. Still, reasons, to be valid, must comply with certain logical requirements. For example, the truth of A is logically a sufficient reason for upholding "A or B"—but the cause for making the latter claim may be the need or wish to rope B into the argument. Hence, although neurophysiology is ultimately to describe and explain reasoning, it cannot justify it. (Analogy: sociology can explain crime but not condone it.) Only logic can justify reasoning by showing that it fits accepted canons of inference.

The most basic inference pattern is substitution, which is subject to the principle of (invariance under) substitution of variables (terms) or of entire propositions. If in any formula $F(\ldots x \ldots)$ the variable y is substituted for x, then that formula implies $F(\ldots y \ldots)$. For example, $x^2 + x = 0$ and $x = y - 1$ jointly imply $y(y-1) = 0$. The most familiar method for solving systems of linear equations, and the most popular method for evaluating integrals, use the rule of substitution.

The simplest of all the rules of inference involving two premises is of course the *modus ponens*: B is an immediate consequence of A and $A \Rightarrow B$. Combining this rule with the rule of substitution (exchanging A for $\neg B$ and B for $\neg A$) yields the *modus tollens*: $\neg B$ and $A \Rightarrow B$ jointly entail $\neg A$. Many a scientific paper violates either rule. Indeed, a rather common fallacy is this: $A \Rightarrow B$ and B, hence A (or $A \Rightarrow B$ and $\neg A$, hence $\neg B$). This error is in fact committed when a hypothesis is declared true because some of its consequences have been confirmed.

Needless to say, the detectives in good mystery stories do not commit such fallacies. A typical example is this. After some inquiries the inspector gathers all the people involved in the crime and presents them his premises:

1. Every criminal has a motive and an opportunity. (Law of criminology.)
2. The butler had both a motive and an opportunity. (Empirical datum.)

3. Nobody except the butler had both motive and opportunity. (Plausible hypothesis.)

All three premises are necessary and sufficient to draw the conclusion that the butler did it. This proof can be obtained by *reductio ad absurdum*: Suppose someone else, call him or her C, did it. Then C had both motive and opportunity. But this contradicts premise No. 3. Hence C is identical with the butler. The argument is an example of valid *hypothetical* reasoning, for it involves two hypotheses going beyond the hard datum 2, namely 1 and 3.

To be sure, we are not always given conditionals: sometimes our task is to prove them. In mathematics, to prove that A implies B one can often assume A and check whether it entails B (i.e. whether "$A \Rightarrow B$" is a tautology). For, if it does, then a fortiori $A \Rightarrow B$. (A second method is to assume $\neg B$ and find out what this entails concerning A. If it entails that $\neg A$, the conditional has been established.) In factual science and technology the method for establishing $A \Rightarrow B$ is quite different. One enforces A and checks whether B obtains. If it does, the conditional is confirmed but not proved, for it may well happen that only the favorable cases were observed during the given experimental run. If B does not obtain, the conditional has been refuted—or so it seems. In general, there are no rigorous proofs of factual statements.

Strictly speaking, mathematical proofs are not absolutely rigorous either, and this simply because there are no absolute (definitive) standards of rigor. Given any proof, it is likely that it can be improved on, if not right now, then in a few years or centuries. Moreover, logical standards are changeable. The history of the calculus is littered with proofs that were regarded perfect at one time but turned out to be defective in some respect or other. However, there is a big difference between mathematical proof and factual confirmation: whereas the former can *in principle* be rendered unassailable relative to the prevailing standards of rigor, factual confirmation remains in *principle* assailable, not only because its premises are uncertain but also because it involves plausible rather than demonstrative inferences. Caution: the fallibility of a mathematical proof and of a factual confirmation does not entail that we must ultimately resort to some extrinsic sanction such as consensus. It only suggests that further research may be needed.

Our discussion of deductive inference has not depended upon the notion of truth. To be sure, we did characterize a rule of deductive inference as a truth-preserving function from n-tuples of propositions to propositions. However, this condition is needed to check the validity of any proposed rule of inference, not to perform any deductions. (Thus the obviously false premises "All men are immortal" and "Donald Duck is a man" entail rigorously "Donald Duck is immortal".) It is just as well that deducibility should not involve truth, for often (*a*) we do not know whether our premises are true (but are just trying them out), and (*b*) our premises are at best half-true.

Logic, the theory of deduction, is concerned to save consistency, not to guarantee truth. We want consistency for the following reasons: (*a*) a contradiction entails any proposition, and is thus a source of unwanted proliferation (conceptual cancer); (*b*) consistency is necessary (though not sufficient) for a set of propositions to form a system allowing one to jump safely from one component to another. Logic (or rather metalogic) can diagnose consistency and inconsistency but is impotent to advise what to do in the face of inconsistency. To be sure, if a set of premises turns out to be inconsistent then we can strike off some premises to obtain one or more contradiction-free subsets of premises, but logic will not tell us which, if any, such subsets is formed by true premises.

For example, the set of premises $\{A, B, A \Rightarrow \neg B\}$ is inconsistent. In fact, the 1st and 3rd imply $\neg B$, which contradicts the 2nd; and the 2nd and 3rd imply $\neg A$, which contradicts the first. The three minimal surgeries yielding consistent sets of premises are:

(i) Drop A, leaving $\{B, A \Rightarrow \neg B\}$, which reduces to $\{\neg A, B\}$.
(ii) Drop B, leaving $\{A, A \Rightarrow \neg B\}$, which reduces to $\{A, \neg B\}$.
(iii) Drop $A \Rightarrow \neg B$, leaving $\{A, B\}$.

In the case of mathematics all three subsets may be correct though of unequal or even null interest. If we are to choose among them it will be on grounds of richness or beauty, economy or applicability. If the premises are factual, compatibility with the available evidence and with other theories may indicate which of the three subsets, if any, is the best (truest and deepest) set. But it may well happen that all three are inadequate—or that the evidence is so hazy that neither is completely ruled out by it. In conclusion, *pace* Rescher (1976), there is no a priori general theory of plausibility that could help us choose among the above contradiction-free subsets.

Consistency is an absolute desideratum relative to each particular body of knowledge, in particular theory. It is not equally desirable with regard to the totality of knowledge at a given moment: if overall consistency were required at every moment, then no rival views would be tolerated, and consequently we would be unable to correct any errors. So, inconsistency is a fact of life and we must learn to eliminate it (by uprooting errors) rather than to avoid it.

Dialectical philosophers hold that contradiction is not only unavoidable but also desirable to keep things moving, for contradiction among ideas (and even among facts) is the source of all change. The evidence goes against the universality attributed to the genetic power of contradiction. Indeed the history of ideas shows that, whereas some conceptual contradictions have indeed promoted progress (by calling for strenuous efforts to surmount them), others have resulted in stagnation or even regress. Here is a random sample of counterexamples to the dialectical thesis. (*a*) The contradiction, within the Pythagorean school, between its tenets concerning rational numbers and the discovery of irrationals (such as the square root of 2), brought the efforts of the members of the school to a halt. (*b*) The contradiction between the subjectivist and the frequency interpretations of probability has produced no new results except for confusion. (*c*) The contradiction between the wave and the particle interpretations of the quantum theory has held back work in the foundations of the latter. (*d*) The contradiction between classical mentalistic psychology and behaviorism kept psychology from making any breakthroughs for decades. (*e*) The contradiction between individualism and holism in social science is still bogging it down. The moral is clear: Contradictions, when protracted, are unhealthy because they generate sterile and acrimonious debates rather than new ideas or new experiments. Contradiction does not generate anything: at most it may motivate some people to overcome it.

We often need to estimate the degree of agreement or disagreement between two sets of propositions, e.g. between two scientific theories, or two ideologies, or a scientific theory and an ideology. Actually we need two different measures, one for the degree of agreement between two sets of propositions, and another for the degree of compatibility. The two measures must be radically different because two sets of propositions may be quite different yet compatible (e.g. an atomic theory and a sociological theory), or similar yet mutually incompatible (e.g. two rival theories).

(Consider two sets of propositions $P=\{P_i | 1 \leq i \leq p\}$ and $Q=\{q_j | 1 \leq j \leq q\}$. We define the *degree of agreement* between P and Q as the fraction of the number of propositions they share, i.e.

$$A(P,Q)=(1/p+q)\cdot|P \cap Q|.$$

The degree of *disagreement* between P and Q is defined as

$D(P, Q) = 1 - A(P, Q)$. The values of A and D are comprised between 0 and 1. Such values can be computed only for finite sets of propositions. In the case of theories we may take the sets of their postulates; in the case of rival theories it pays to enrich the postulates with all the standard theorems. Two different presentations or organizations of the same theoretical material are supposed to agree fully with one another, i.e. $A(P, Q) = 1$. On the other hand two genuinely rival theories must share some but not all their propositions, i.e. $0 < A(P, Q) < 1$.)

(Let us now introduce the concept of degree of compatibility between two sets P and Q of propositions. They will be totally incompatible if every member of P contradicts each component of Q—an extremely unlikely case. We abbreviate 'p_i is incompatible with q_j' to '$p_i | q_j$', and correspondingly 'p_i is compatible with q_j' to '$\neg(p_i | q_j)$'. We now define the *degree of compatibility* between P and Q as the fraction of the number of pairs of compatible propositions:

$$C(P, Q) = (1/pq)|\{\langle p_i, q_j \rangle \in P \times Q \mid \neg(p_i | q_j)\}|.$$

The dual is of course the degree of *incompatibility* between P and Q, namely $I(P, Q) = 1 - C(P, Q)$. The values of the two measures are comprised between 0 and 1. If $C(P, Q) = 1$, P and Q are fully compatible, even if they do not agree at all, whereas they are totally incompatible if $C(P, Q) = 0$. Any value between these extremes indicates incomplete compatibility.)

Can logic help us obtain new knowledge? Nearly all of the 17th and 18th century philosophers derided logic for its alleged impotence to do so. They demanded instead that a logic of invention and discovery be instituted. Although such a logic was never born, it was much praised. And in the meantime the contempt for logic blocked its advancement. (You do not farm well a land that you despise or hate.) Logical investigations were resumed two centuries later by mathematicians, but the calumny that logic is useless to advance knowledge still remains. Even after mathematical logic was well established, a distinguished philosopher such as Schlick (1925) ignored its role in research and suggested that, since science studies only facts, and there are no negative facts, it should make do with a negation-free logic. This is a correct inference from the thesis that scientific research is just the collecting of trivial facts. It falls down as soon as one realizes that there is no science without conjecturing and arguing, which involves counterconjecturing and counterarguing. Certainly there are no negative facts, but negation (a "positive" brain process) is part of every inquiry process. Statements of the form "A is not B" are just as important as those of the form "*A* is *B*"; i.e. identity and inclusion are not more important than difference and exclusion.

When a mathematician, a theoretical scientist or a technologist prove a new theorem, they produce genuine new knowledge with the help of the logical theory underlying (presupposed by) the theory or theories they employ. (And, since the most common types of proof are by *reductio ad absurdum* and by counterexample, negation is usually involved in such deductive processes.) To be sure, logic does not supply any substantive premises: it acts as midwife rather than parent. Nevertheless its intervention is necessary to produce or control the new results. If deduction did not yield new knowledge we would not care to perform it and, in particular, we would not care to build or use theories proper, all of which are hypothetical-deductive systems.

It is often rejoined that the conclusion of a deductive inference is "contained" or "included" in the premises, so it is not actually new. This is sophistry. In the first place not all the conclusions of a deductive inference occur among the premises. Thus "All *A* are *C*" follows rigorously from "All *A* and *B*" and "All *B* are *C*" (as well as from "All *A* are *X*" and "All *X* are *C*"), but is not contained in these premises. Likewise one gains new knowledge, by deduction, from the data that $a > b$ and $b > c$, namely $a > c$.

Secondly, even if all deduction were merely unpacking, it still would supply new knowledge: if nobody knew the conclusion before it was drawn, then the conclusion was not an item of knowledge—it only became one. What is true is that we may *feign* that the conclusion exists (formally, not epistemically) the moment the premises (and the underlying logic) are laid down. Likewise the value of a continuous function at a given point is "given" by a certain formula containing the values of the function and all its derivatives at some other point, but the former is unknown as long as the computations indicated by the formula have not been performed.

In short, deduction is a prime tool of inquiry, for it allows us to explore the (a priori unknown) consequences of our assumptions and definitions. And, because deductive reasoning is formalizable and subject to explicit rules, one is justified in calling *hypothetico-deductive method* the pattern of reasoning consisting in assuming, deducing, and checking. No other type of inference is formalizable and subject to explicit rules, and therefore capable of becoming the nucleus of a method. In particular, although we all use plenty of intuition, analogizing and induction, there is no such thing as the *method* of intuition, or analogy, or induction. There are only philosophical tracts on such ghostly subjects.

Finally, are there things we cannot reason about? All we know is that, every so often, we come up against certain limits to knowledge that are subsequently overcome. Thus the Eleatic philosophers believed change to be contradictory and therefore contrary to reason; and, being rationalists, they declared change to be illusory. Thus a moving arrow was said *to be* and *not to be* at a given place at a given time. Hegel kept this archaic mode of thought and admitted the Eleatic thesis that change is contradictory, but claimed that this only calls for the dialectical mode of thought. Bergson chose the other horn of the ancient dilemma: he claimed that only intuition can grasp change—whereas science, which is rational, cannot. We see no contradiction in change. Instead of saying that the arrow is, and is not, at a given place at a given time, we say that it moves with such and such instantaneous velocity through a given place. We replace being with becoming as the basic category, and use the differential calculus to reason about change. One man's limits to reason is another's challenge to improve reasoning.

2. PLAUSIBLE REASONING

2.1. *Analogical Inference*

The most basic judgment we can form about a collection of items is that they are, or fail to be, similar in some respect(s)—i.e. that they are analogous or disanalogous. Judgments of analogy underlie a number of judgments of other types, in particular collecting or grouping (Ch. 5, Section 1.2) and inductive generalization (Section 2.2). We make them all the time in daily life and in all fields of inquiry. Here is a tiny random sample of analogies that have been fruitful at one time although they ended up by blocking further research: set–collection of things, gas–solute, electric current–fluid current, light–sound, economic competition–natural selection, optics–mechanics, and rumor propagation–epidemics. We tend to remember only the fruitful analogies. The list of misleading analogies, if it could be drawn, is likely to be more impressive.

Analogical inference is based on analogical judgment but goes beyond the latter: from the fact or assumption that two objects are analogous in some known respect(s) it "concludes" that they *may* be analogous in some other unknown respect(s) as well. Thus organicists, likening society to the human body, conclude that the workers, being the muscles of society, ought to obey the brain of society, namely its ruling class. And Copernicus, likening the solar system to a monarchy, argued that, just as the monarch is surrounded by his vassals and courtiers, so the sun must be at the center of the solar system. Both are of course implausible reasonings, not plausible ones. They are faulty because they rest on wrong analogy judgments: society is not like an organism, and the solar system is not like a monarchy. On the other hand Aristotle's classical example of analogical reasoning, namely "Gills are to water what lungs are to air", was plausible and fruitful. It was so because it rested on two genuine analogies: gill—lung and water—air. A necessary condition for a plausible and fruitful analogical inference is then that it rests on genuine analogies. We should therefore elucidate the rather vague notion of analogy.

There are several ways of exactifying the notion of analogy, and each is suitable for a given set of problems. One of them involves the modeling of the analogous objects as sets. In this case one can elucidate a number of analogy strengths, from mere some–some correspondence to homomorphism, to isomorphism or perfect similarity. (Cf. Bunge, 1973b, Ch. 6.) A second method is to count the number of properties shared by the analogous things: see Vol. 3, Ch. 2. A third applies when the things concerned are systems, i.e. objects with a definite composition, environment, and structure. We proceed to elucidate this notion (Bunge, 1981b).

Consider two systems, each modeled qualitatively by a triple *composition–environment–structure* (Vol. 4, Ch. 1). Each component of this triple is a set. In comparing the triples representing two systems we compare sets (of components, environmental items, or relations). Therefore we must start with the concept of similarity or analogy of two sets. We stipulate that two sets are similar with respect to a third set if they overlap with the latter. More precisely, if A, B and C are nonempty sets, then ***A is similar to B with respect to C*** if, and only if, $A \cap C \neq \varnothing$ and $B \cap C \neq \varnothing$. We can now define the concept we want:

Let s_1 and s_2 be two concrete systems with known compositions, environments, and structures. Then:

(i) s_1 and s_2 are *substantially analogous* iff their compositions are similar, i.e. if there is a natural class or species that has a nonempty overlap with those compositions;

(ii) s_1 and s_2 are *environmentally analogous* iff their environments are similar, i.e. if there is a natural class or species that has a nonempty overlap with those environments;

(iii) s_1 and s_2 are *structurally* (or *formally*) *analogous* iff their structures are similar, i.e. if there is a family of relations that has a nonempty overlap with those structures. (In particular, two systems are *nomologically analogous* if they share some laws.)

All organisms are substantially analogous, if only because they all contain DNA molecules. Fish and marine mammals are environmentally analogous. The human hand and the bat wing are structurally analogous (homologous).

The above concepts may be quantitated with the help of the following concept of degree of similarity between sets. If A and B are sets intersecting a third set C, then A and B can be said to be fully analogous with respect to C if both have the same number of elements of kind C, i.e. if the cardinalities of $A \cap C$ and $B \cap C$ are the same. The degree of dissimilarity between A and B with respect to C is then measured by the difference between $|A \cap C|$ and $|B \cap C|$. More precisely, we stipulate that

$$d_C(A, B) = \frac{||A \cap C| - |B \cap C||}{\max\{|A \cap C|, |B \cap C|\}}.$$

Therefore the *degree of similarity between A and B with respect to C* may be taken to equal the complement of $d_C(A, B)$ to unity, i.e. $s_C(A, B) = 1 - d_C(A, B)$. The values of the degree of relative similarity are comprised between 0 and 1. If A and B contain no C's, then $s_C(A, B) = 0$. But if there are just as many C's in A as in B, then $s_C(A, B) = 1$ even though they are not the same individuals. For two sets to be similar in the given respect C they just have to contain at least one member of kind C.)

By applying this concept of similarity between sets to our previous notions of systems analogy we obtain the concepts of degree of substantial, environmental, and structural analogy. Moreover, by adding all three and dividing by 3 we get the concept of *degree of total analogy*. Finally, we can invert the process to obtain the qualitative concepts of analogy in terms of the quantitative ones: Two systems are (*a*) *analogous* iff their degree of total analogy is greater than 0; (*b*) *weakly analogous* iff their degree of total analogy, though nonvanishing, is close to 0; (*c*) *strongly analogous* iff their degree of total analogy is close to 1.

A proposition of the form "A and B are analogous" is *true* if, in fact, the degree of analogy between A and B is greater than 0. However, truth is not enough, for it may be deep or shallow, fruitful or barren. The judgment of analogy will be *superficial* if the systems in question are weakly analogous. (*Example*: the brain–computer analogy.) The judgment will be *deep* if the systems in question are strongly analogous. (*Example*: the man–ape analogy.) However, even deep analogies may not be fruitful. The fruitfulness of an analogy depends not only upon its degree but also on the problem at hand and the brain that uses it. Many people had noted the analogy between man and other primates; in the Middle Ages it was a common belief that monkey was a fallen man. However, it took a Charles Darwin to use this analogy to place man and ape in an evolutionary perspective.

Because the fruitfulness of an analogical inference depends on the actual degree of analogy as well as on the imagination of the user, there can hardly be a general theory of analogical reasoning. The most we can do is to formulate some useful methodological (not logical) rules, such as:

R1. Use only deep analogies—i.e. make sure that the degree of (substantial, environmental, structural, and if possible total) analogy is significantly greater than 0.

R2. Regard all analogies as heuristic devices that may have to be discarded eventually.

The first rule is obvious: weak analogies lead nowhere or, worse, they may block inquiry. For example, the analogies between brain and hydraulic system, electrical network, telephone exchange, or computer, are not only extremely weak: they have also misled uncounted workers in neuroscience and in psychology into studying technological gadgets rather than living brains. (More in Vol. 6.)

The second rule is justified as well: even if initially fruitful, an analogy is no substitute for an investigation of the peculiarities of the thing of interest. Thus the quantum-classical analogies were initially useful as rough guides, but ended up by obscuring the originality of the quantum theories as well as the nonclassical nature of their referents (Bunge, 1973a). And the information-theoretic metaphors employed in biochemistry, molecular biology, and genetics, helped crack the genetic "code" but they are now blocking the way to an understanding of the mechanism of protein synthesis and, in general, the way "genetic information" is locked in nucleic acids. (More in Vol. 6.)

Because rigorous reasoning is nonanalogical, scientific papers in advanced science and technology are not supposed to resort to analogical reasoning—not even when proposing analogies. If they do, they do so inadvertently. On the other hand analogy and analogical inference are rampant in the popularization literature, for they are (wrongly) assumed to convey technical ideas to the lay. (Actually they skirt the problem of conveying them.) And, since most philosophers have no access to the original scientific or technical literature, they sometimes get the impression that analogy, far from being just a heuristic trick and a pedagogical prop, is of the essence of science and technology. (For the quaint thesis that scientific theories are analogies or metaphors, see Black (1962) and Hesse (1966).) Occasionally there are also more serious motivations for this view, or analogism. One is fictionism, or the doctrine that we cannot know what and how things really *are* but at most what they are *like*, so we should present our hypotheses and theories as stating only that things behave or appear *as if* they were such and such. (Fictionism has its roots in Kant, was sketched by Nietzsche, and worked out by Vaihinger (1920).) Another root of analogism is the attempt to present science as no better than religion: just as (according to Aquinas) we can know God only by analogy with worldly entities, in particular ourselves, so science would know the world only by analogy with our experience.

In short, analogies and analogical inferences are indispensable tools of inquiry. But analogies can be superficial or deep, and analogical inferences can be misleading or fruitful. It is therefore necessary to assess the respect and degree in which an analogy holds, and to check the conclusion of an analogical argument. And we must not conclude to analogism from the fact that analogy is pervasive, if only because it is not more so than other categories.

2.2. *Statistical Inference and Hypothetical Reasoning*

Induction is, of course, generalization from known (or partially known) instances. Every higher vertebrate performs inductive inferences, and scientists and technologists are no exception. Many of the generalizations, true and false, we form and use in daily life are inductions. (Perhaps most of them are generated by only a handful of cases.) And every time we evaluate the performance of hypotheses, theories, methods and plans, we formulate inductive judgments. So, induction occurs sometimes in the inception of general constructs and always in their validation.

We distinguish the following types of *inductive inference.*

(i) *First degree induction*: from instances to low level generalization. *Pattern*: "All A's up to the nth were found to be B's. Hence it is likely that all A's are B's." *Example*: All the measured values of the speed of light in a vacuum are the same to within experimental error. Hence it is likely that all light waves in a vacuum have the same speed.

(ii) *Second degree induction*: from low level generalizations to higher level generalization. *Pattern*: "Law L holds for every member of a family of sets up to the *n*th. Hence it is likely that law *L* holds for every member of the family." *Example*: The theory of evolution holds for all studied species. Hence it is likely to hold for all biospecies.

(iii) *Statistical specification*: from population to sample. *Pattern*: "The observed fraction of *P*'s in the population *U* is *p*, and *S* is a random sample of *U*. Hence it is likely that the fraction of *P*'s in *S* be near *p*." *Example*: The last census showed that 10 percent of the population are unemployed. Hence it is likely that about 10 percent of the population of Medianville, which is pretty representative, are unemployed.

(iv) *Statistical generalization*: from sample to population. *Pattern*: "*S* is a random sample of the population *U*, and the observed fraction of *P*'s in the random sample *S* of *U* is *p*. Hence it is likely that the fraction of *P*'s in *U* be near *p*." *Example*: In a random sample of college students amounting to the one-hundredth part of the total student body 1 percent were found to be original thinkers or doers. Hence it is likely that the corresponding percentage in the whole student population be about 1 percent.

Note the following traits common to all of these inference patterns. First, neither is wild: they all proceed from supposedly sound (checked) premises. Second, neither leads to new concepts; in particular, all the theoretical concepts occurring in the conclusions appear also in the premises. Third, neither is conclusive or demonstrative: they are not of the form "*A*, hence *B*", but rather "*A*, hence it is likely that *B*", with no specification of the degree of likelihood. Thus room is made for failure of the inferential process. For instance, in the example of (iii) Medianville, though representative on the whole, may be particularly lucky, or unlucky, at the moment the inference is drawn—e.g. because it happens to host an industry that has suddenly boomed or slumped. And in the example of (iv) the student sample, though representative on the whole, may be exceptionally rich or poor in talented students at the time the sample is drawn—e.g. because talented students tend to concentrate in certain colleges.

Consequently neither of the above inference patterns can be erected into a rule or principle, e.g. the "principle of induction", more often praised or criticized than formulated. The conclusions of all such plausible reasonings are just as hypothetical as their premises. Consequently they call for further testing. Besides, neither the premises nor the conclusions occurring in inductive inference are enough: we also need higher level hypotheses, which cannot be generated by studying particulars, if only because they contain concepts that are far removed from observation. The task of induction is not to generate such high level hypotheses in step by step processes, but to take part in their empirical tests, as will be seen in Ch. 12.

In addition to the above four patterns of inductive inference we shall mention four kinds of *hypothetical reasoning*.

(i) *Weak modus ponens*: weak detachment of consequent upon weak assertion of conditional or of antecedent. *Patterns*: (*a*) "If p then q, and 'p' is verisimilar. Hence 'q' is verisimilar"; (*b*) "'If p then q' is verisimilar, and p. Hence 'q' is verisimilar." (*c*) "'If p then q' is verisimilar, and 'p' is verisimilar. Hence 'q' is verisimilar." *Example*: Heavy smoking is likely to cause lung cancer. John is a heavy smoker. Hence John is likely to catch lung cancer."

(ii) *Weak modus tollens*: weak rejection of antecedent upon strong or weak assertion of conditional and weak or strong negation of consequent. *Patterns*: (*a*) "If p then q, and '$\neg q$' is verisimilar. Hence '$\neg p$' is verisimilar". (*b*) "'If p then q' is verisimilar, and $\neg q$. Hence, '$\neg p$' is verisimilar." (*c*) "'*If* p then q' is verisimilar, and '$\neg q$' is verisimilar. Hence '$\neg p$' is verisimilar." *Example*: It is likely that philosophy flourishes only in freedom. Now, there is no freedom in Dogmaland. Hence it is unlikely that philosophy flourishes in Dogmaland.

(iii) *Strong reduction*: weak assertion of antecedent upon assertion of consequent. *Pattern*: "If p then q, and q. Hence maybe 'p' is true." *Example*: brain injuries cause intellectual deficits. Now, Peter has recently converted to a mystical doctrine. Hence it is possible that Peter suffered a brain injury.

(iv) *Weak reduction*: weak assertion of antecedent upon assertion of consequent and strong or weak assertion of conditional. *Patterns*: (*a*) "If p then q, and 'q' is verisimilar. Hence it is possible that 'p' is true." (*b*) "'If p then q' is verisimilar, and q. Hence it is possible that 'p' is true." (*c*) "'If p then q' is verisimilar, and 'q' is verisimilar. Hence it is possible that 'p' is true." *Example*: Famines are likely to cause political unrest. Now, there was a revolution in *A* during period *B*. Hence it is possible that there was a famine in *A* during *B*.

Like inductive reasoning, plausible hypothetical reasoning is educated rather than wild, as well as inconclusive: its conclusion is at best another conjecture worth being investigated. But, unlike inductive reasoning, plausible hypothetical reasoning may proceed from premises containing transempirical concepts. For this reason it occurs not only in the early stages of empirical research and in the testing stage but in all stages of scientific and technological inquiry.

Inferring from symptoms, or indicators, is a common species of hypothetical reasoning and, in particular, of weak reduction. Thus since all infections cause an abnormal count of white blood corpuscles, we tend to attribute the latter to some infection; however, it may have a different cause, such as cancer of the bone marrow. Likewise we tend to attribute criminality and drug addiction to poverty, but some times it is caused by alienation.

Most indicators in the life sciences and in the social sciences are ambiguous, i.e. they occur in hypotheses of the form *If A or B then C*, where '*C*' denotes the observed variable adopted as an indicator of *A* or *B*. Therefore it is advisable to employ a whole battery of indicators in order to narrow down the set of possible causes. But the only effective way of removing the uncertainties in inferences from symptoms is to employ functional relationships between symptoms and causes. Thus the barometer gives an unambiguous reading of atmospheric pressure because it is built on the law of the linear relation between atmospheric pressure and barometer height; and it can be used to give an unambiguous value of the altitude on the strength of the exponential law relating altitude and atmospheric pressure.

Plausible reasoning, whether analogical, inductive, hypothetical, or of some other kind, cannot be regimented any more than intuition or insight can. First, because it depends essentially on the nature of the case. Second, because it depends on the talent of the subject to notice or conjecture relevant features or patterns. Third, because one and the same set of data is compatible with any number of alternative hypotheses: Figure 6.1. In sum there can be no *logic* of plausible reasoning.

However, plausible reasoning can be controlled. Actually it is being controlled by mathematical statistics, not by philosophy. Indeed, the tasks

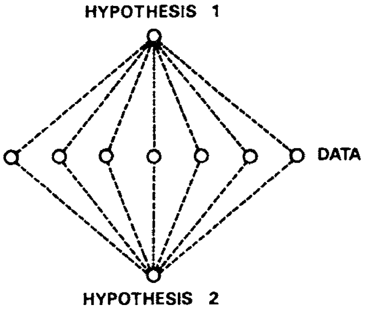


Fig. 6.1. At least two different (rival) hypotheses are compatible with any given set of data.

of that discipline are not only to help collect, screen and organize data, but also to help compare the latter with hypotheses in order to evaluate both (usually the hypotheses). But, of course, mathematical statistics produces neither raw data nor the hypotheses accounting for them: it is a device for control not generation. For example, it tells us that statistical generalization and specification are plausible only if the sample is random and its size not too small: otherwise violent fluctuations around the frequency will be observed. (Sample size, though essential, is less so than randomness.)

We all pay lip service to mathematical statistics but few of us succeed in avoiding committing statistical fallacies such as the gambler's fallacy. This is the belief that every segment of a random sequence (or every random sample of a population) must have the same properties as the entire sequence (or population). Thus if in coin tossing the inexperienced gambler gets five heads in a row, he may be inclined to bet that the next flipping will produce a tail: he believes that there must exist a force producing a deviation that will cancel the previous deviation from the long run 0.5 frequency. This fallacy is committed not only by gamblers but also by a disquieting large fraction of scientific researchers (particularly in psychology and social science); moreover, they exhibit undue confidence in early trends, overestimate significance, and offer causal explanations for any local discrepancies instead of paying attention to sample variability (Tversky and Kahneman, 1971). This finding shows how inadequate our statistical education is, and it suggests that the teaching of probability and statistics should start earlier than usual, perhaps in primary school.

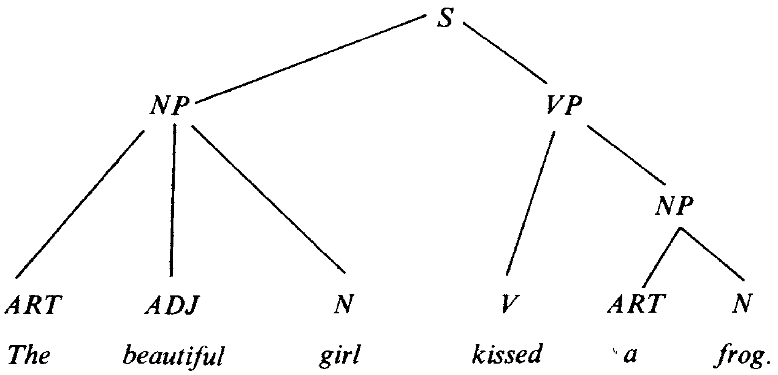
3. Discussing

3.1. *Analysis and Criticism*

Can we learn from discussing ideas, e.g. from analyzing them and arguing about them? Rationalists reply in the affirmative, and some of them claim that analysis, criticism and argument are the very best cognitive methods. The schoolmen of all times assent because they hope to extract all the knowledge locked in allegedly canonical and infallible texts. The linguistic philosophers, because they regard ordinary language as the repository of wisdom. The dialectical philosophers, because they believe that new ideas can be generated only by the clash of old ideas (or the clash between ideas and facts). And the critical rationalists, because they are interested primarily in the weeding out of false ideas. On the other hand, empiricists deny all this: to them only experience, in particular observation, can give us new knowledge, whereas discussion is unprofitable and perhaps even (as Comte said) frivolous. Who is right? Once again it will turn out that each of the contending parties holds a grain of truth.

To analyze a complex object, such as a system, is to exhibit its components, environment (or context), and structure (or organization). The analysis can be partial (decomposition into subsystems) or total (decomposition into ultimate or elementary components). It can be empirical (e.g. chemical) or conceptual (e.g. logical). In every case the components, the (relevant) environmental items, and the relations or links holding the system together are identified and classed. A few examples will show the gist and scope of such procedure.

Example 1. Structural analysis of a sentence. Look at the tree diagram representing the surface structure of the English sentence *The beautiful girl kissed a frog.*



The species or lexical categories are ART (icle), ADJ(ective) and N(oun); the genera NP (noun phrase) and VP (verb phrase). One and the same object, namely the sentence S, is conceived of as a system and analyzed on two levels: subsystems (made up of objects on the genus level, NP or VP), and elementary objects (belonging to species such as N and V). The analysis uses then a word systematics that is taken for granted, and which differs somewhat from language to language. (Thus Latin lacks the ART species.)

Example 2. Logical analysis of the proposition expressed by the sentence in Example 1. A possible analysis is this:

For some x, x is a frog and b kissed x, where

> $b=$ The y such that y is a girl and y is beautiful at fixed place p and time t.

Example 3. Mathematical analysis of proposition in Example 2. Every one of the logical categories occurring in the given proposition can be analyzed further, i.e. into deeper concepts. For example, the predicate "frog" can be construed as a function from the set of all reptiles (or all vertebrates, or all animals, or all organisms) to the set of all propositions contining that predicate. Likewise the predicate "beautiful" can be analyzed as a function from the cartesian product of the set of all objects, the set of all places, and the set of all times, to the set of all propositions containing that predicate. The connective "and" can be analyzed as a function from pairs of propositions to propositions. The particularizer "Some" can be analyzed as "Not-for-all-not", where "all" and "not" are primitives, or in a more sophisticated way. Likewise the descriptor "the y such that" can be analyzed *à la* Russell or as in Vol. 2, Ch. 9.

In all three cases the "environment" (context) of the system being analyzed has been taken for granted: the English language in the first, the predicate calculus in the second, and mathematics in the third. And every analysis has presupposed some body of knowledge about such contexts—in particular a knowledge of basic categories such as "noun", "predicate", and "function". The logical analysis (Example 2) has also presupposed a bit of factual knowledge, which on closer examination turns out to be dubious. In fact, that analysis presupposes that beauty is absolute (independent of the beholder); it also presupposes that place and time are absolute (independent of the reference frame). If relativized, the predicate turns out to be of degree 5 or higher, not 3.

In performing each of the three analyses we had to invest a certain amount of knowledge. But we gained some new knowledge too. For, even supposing that we knew the components of the system before analyzing it (since we fashioned it ourselves), we did not know the kind or type to which each of them belonged, or how they were related to one another. To be sure, the output has been puny by comparison with the input. However, the same can be said of all kinds of analysis. Think of the analysis of a function as an infinite series, or the analysis of a chemical compound into atoms, or the analysis of a society into economic, political, and cultural subsystems. In all these cases the knowledge investment (data and general principles) is far greater than the knowledge produced. But what is important is that *(a)* analysis *can* produce new knowledge—empiricism notwithstanding, and (*b*) analysis is only *one* of the ways of gaining new knowledge from available knowledge—rationalism notwithstanding.

Analysis can be valuable, but not all analysis is. Thus linguistic analysis, as practised by Wittgenstein and his followers, and hermeneutic analysis, the daily fare of lawyers and popular among theologians, political ideologists, and contemporary German philosophers, are ambivalent. The former because it is shallow and sometimes misleading: it makes no use of logic—let alone mathematics or science—and is limited to examining the use of ordinary language locutions. Its results are usually very modest, as must be expected from a "philosophy without tears" (Russell). And sometimes they are grotesque, as when Ryle (1954) concluded, from an examination of what he called 'the grammar of the word *perceiving*', that seeing and hearing are not states or processes, hence not something to be located, inspected and measured. Correct conceptual analysis calls for more than a mastering of one's dialect: it requires substantive knowledge and fine analytic tools, often mathematical ones.

As for hermeneutics, or the interpretation of texts, it has three uses. One is to arrive at an understanding of the original text, another is to evaluate current actions or ideas in the light of some code, and a third is to help us understand or modify current events. In the first case, the hermeneutic analyst attempts to explain the given text in contemporary terms and to place it in its original cultural context. This enterprise is as risky as it is useful, for it forces the authors of the past to think and write like ourselves.

A second use of hermeneutic analysis is to examine certain actions or ideas to find out whether they conform to some doctrine or code. It is standard procedure in law. The basic inference pattern is this: *A holds or does B. Now, B is compatible (or incompatible) with canonical text C. Hence A is right (or wrong) on C.* The doctrine or code itself is not questioned. So, the analysis is useful if the canonical text happens to be correct or just, but it can be noxious if the text is wrong or unjust. In any case this kind of hermeneutic analysis does supply some knowledge—e.g. that A did say or do something fitting (or clashing with) a certain canonical text. Moreover, if what A said or did is judged as correct or just on a different standard, then that piece of hermeneutic analysis will suggest that the canonical text is wrong on the given point, which in turn may elicit its reform.

Finally, a third kind of hermeneutic analysis is to try and understand what happens to and around us in terms of some text written hundreds or even thousands of years ago for different people in different circumstances. Such analysis is worthless or worse, as it involves a distortion of the text, the contemporary events, or both—for otherwise the events could not possibly fit the text. Needless to say, this use of hermeneutic analysis is common among ideologists wishing to confirm their beliefs or to use them to praise or condemn some of their contemporaries. Arguing with analysts of this kind is unlikely to produce new knowledge, for their intention is apologetic, not inquiring.

Conceptual analysis can be regarded as a means for uncovering composition, environment or structure, or as part of criticism. And criticism is of two kinds: *apologetic* (or in defense of a belief system) or *inquiring* (or aiming at spotting error). In turn, inquiring criticism can be *negative* or *constructive*. The former stops when error has been found and exposed as such. Constructive criticism goes farther, to error correction. The proofreader, the dedicated teacher and the productive investigator, technologist, and manager are constructive critics. On the other hand the literary critic and the art critic, as well as the typical book reviewer, are negative critics. Negative criticism is easier and therefore more common than constructive criticism: (*a*) it bears on something that has already been set up (by oneself or someone else), and (*b*) the denial of a proposition has better chances of being true than its assertion. (*Example*: if one states that the current world population is exactly 4 billion, he shall be in error. The denial of that statement is true—a cheap truth, for it gives no positive information.)

Criticism is necessary in all fields of inquiry because an aim of all inquiry is truth, and no research is free from error. We are bound to make some erroneous assumptions, observations, computations, or inferences; we are bound to employ vague concepts or inadequate methods; and we are bound to disregard some variables and overrate others. Any such error may spoil a research project if it goes undetected and uncorrected, but none of them counts in the end if spotted and repaired in time. This is why criticism is tolerated, nay encouraged, in science, technology, the humanities, and enlightened management and politics. For the same reason criticism is discouraged elsewhere, particularly in institutions devoted to the maintenance of dogma, authority, or privilege. Thus churches meet criticism with excommunication, the armed forces with court martialling, and some political parties with expulsion.

The epistemology of science and technology involves methodological (though not systematic) skepticism, a critical attitude that spares nothing excepts the ideals of objectivity and rationality. Everything, from theories to data, from methods to institutions, can become the object of scientific doubt or technological redesign. Science and technology have institutionalized skepticism: "Organized skepticism involves a latent questioning of certain bases of established routine, authority, vested procedures, and the realm of the 'sacred' generally" (Merton, 1973, p. 264). For this reason science and technology are bound to clash sooner or later with all other institutions, for these usually demand faith and loyalty.

The critical attitude, essential in science, technology and the humanities, seems to be rather uncommon elsewhere, and tolerance of criticism quite recent. There must have been dissenters in all human societies, but they are likely to have been persuaded or suppressed by persons wielding authority. All societies have belief-enforcing mechanisms, and only a few tolerate open criticism, particularly of basic social policies. However, in most cases no repression is required to exact conformity, for most people are ready to believe or do what they are told.

In short, criticism is necessary to clarify ideas and eliminate errors. However, criticism should be neither overdone nor overrated. It should not be overdone because it may kill good ideas in the bud, when they are still hazy and have not shown their worth. This is why scientists and technologists resent hasty criticism: they know that every idea is born incomplete and defective, and it should be allowed to develop before being shot down. And criticism should not be overrated (as Popper (1963) does) as the main moving force of every intellectual development because, before an idea can be analyzed, criticized, or rejected, it must have been thought up by somebody. After all, it is easier to destroy than to construct. Just as the history of art is that of creative artists, not of art critics, so the history of science and technology is that of original scientists and technologists, not of thinkers specializing in criticizing as severely as possible every tentative solution to a problem. The main moving force of the advancement of knowledge is not contradiction or criticism but inventiveness in facing and solving new problems. Criticism is the control of the process of inquiry, not its energy source. Like every other control it must be handled with care, for whereas in some cases criticism uproots error, in others it just gives vent to neophobia: witness the editorial process in learned journals. (Cf. Garcia, 1981.)

Criticism is necessary in all conscious endeavors, whether cognitive or practical, but not all criticisms are equally valuable. The best criticism is constructive: one engages in it not to show off (as is the case in carping) or for sheer *Schadenfreude*, but for the love of truth or usefulness. It is part of the self-correcting mechanism of inquiry, be it in science, technology, or the humanities. (Ideology and pseudoscience lack such a mechanism. If only for this reason it is mistaken and dangerous to hold—as Feyerabend (1975) has done—that science is no better than magic or religion.)

To conclude this section. Criticism is neither the mark of rationality (Popper) nor the life of science (Poincaré), but a component of all inquiry. Therefore, to disregard or stifle criticism is to give free rein to dogma and charlatanry—whereas to exaggerate its importance is to forget that it is a means not an end. The clash of ideas is important to clarify and evaluate them, but not more important than the ideas themselves. And there is no such thing as the critical (or dialectical) *method*, i.e. a set of rules for sparking off clashes of ideas, examining them critically, or evaluating them, let alone for generating new ideas superseding the old ones. There are only a few general principles of argumentation, such as "Stick to the subject", "Do not tolerate self-contradiction", and "Concede defeat when it comes". But we are already trespassing on the next section.

3.2. *Controversy*

Controversy is criticism rendered public, at least within some inquiring community. Publicity has the great advantage that many can join in to contribute to the clarification of issues and the resolution of conflicts. But it also has the disadvantage of inflaming noncognitive passions and thus makes admission of error more difficult.

The history of religion, political ideology, and philosophy, is often presented as an unproductive sequence of indecisive fights over firmly held beliefs. In contrast, that of science and technology is usually presented as a progressive sequence of discoveries and inventions—all of them "positive" contributions—wherein controversy plays no important role. (See Agassi, 1963.) Moreover we are often told that controversy is a mark of vanity or, worse, an indicator of the immaturity of an epistemic field. Thus Kuhn (1970, pp. 6–7): "In a sense [. . .] it is precisely the abandonment of critical discourse that marks the transition to science".

There is much truth in the above. In fact controversy is less frequent, and far less indecisive, in science and technology than elsewhere; and it is less conspicuous in the older branches of science and technology than in the newer ones. (Thus there are far more and more acrimonious controversies in psychology or sociology than in mathematics or chemistry.) Moreover, a sort of preliminary dogmatism is needed in science and technology, where one always starts by taking a certain body of knowledge for granted, never questions all of it, and replaces some of its components only if forced to. Kuhn (1963) has pointed out the positive function of "dogma" in scientific research, holding that, as a consequence, resistance to change is normal and on the whole healthy. However, since the inquirer regards his background knowledge as only a launching pad that may have to be modified, it is incorrect to call it a dogma. His may be called *methodical dogmatism*, the companion of the *methodical skepticism* discussed earlier and the enemy of systematic dogmatism. Indeed the "preconceptions" or even "prejudices" that a scientist or technologist inherits are always on probation and may be modified or discarded altogether, some for having been shown to be erroneous, others for having lost value or interest. All scientific and technological (and also humanistic) knowledge has a built-in obsolescence. On the other hand ideological (in particular religious) dogmas are untouchable, and so are the principles of school philosophy. Those who reject them are branded as heretics and perhaps also punished. On the other hand those who criticize accepted scientific or technological principles or practices may be regarded as unorthodox but, if proved right, are hailed as revolutionaries or benefactors.

Just as excess controversy is a sign of immaturity of an epistemic field, so too little controversy may indicate senility, for every radically new idea is bound to elicit controversy if published. The healthy state of an epistemic field is one where most research is done along traditional (but fertile) lines, some explores radically new (but promising) approaches, and another argues the merits and demerits of either. No special control mechanism is needed to produce or keep such division of labor: most inquirers are naturally inclined to follow a successful tradition, only a few are capable of staging revolutions, and the controversy lovers, so numerous in the humanities, are scarce in science and technology. In fact most scientists shy away from controversy either from timidity or because they believe the positivist legend that only "positive" findings matter. As for technologists, they are even more circumspect though for different reasons: the design of a new artifact may be an industrial or military secret or, once in production, controversy over it could hamper its marketing or use. Not many technologists are as public spirited and courageous as Ralph Nader.

There is controversy in nearly every contemporary research field. Sometimes controversy indicates the inadequacy of, hence dissatisfaction with, some received ideas or practices; at other times it indicates resistance to revolutionary innovations; and at still other times, clash with vested cultural, political, or economic interests. Whatever else it may indicate, controversy shows the interest aroused by the issue; in turn interest is a (weak) indicator of importance. (Nobody except Byzantine philosophers or theologians gets excited over trivia.)

Here is a quick sample of recent controversies in science and technology. (*a*) The controversy over the physical interpretation of the mathematical formalism of the quantum theories has continued unabated since the inception of these theories. The main rivals are the positivist (or Copenhagen) and the realist (or objectivist) interpretation. (*b*) The controversy between the "big bang" hypothesis of the origin of the universe and the steady state theory died out in the mid 1960s. However, a few astronomers do not accept either position and are looking for alternative explanations of the redshift of distant galaxies. (*c*) Although most biologists admit the fact of evolution, not all of them believe that it has been gradual or that it has proceeded exclusively on random mutation and natural selection. (*d*) Depending on which traits are taken to be peculiar to *Homo sapiens*, the origin of modern man is assumed to date back either 50,000 or 300,000 years ago. And whereas some investigators postulate a single line, others conjecture that there have been at least two (one of them the Neanderthal man). (*e*) Psychology is torn by the nature–nurture, the behaviorism–cognitivism, and the mind–body controversies. (*f*) In the social sciences the old battles between individualism and holism, as well as between idealism and materialism, are still reging—and these are only two out of dozens of spirited controversies. (*g*) Technologists are divided over the hardness of technology (advanced, appropriate, or soft) that should be used, over the accountability of technologists to society, etc. (*h*) Military experts debate endlessly over the merits and demerits of intercontinental missiles, neutron bombs, and other weapons.

None of these controversies is likely to be barren provided they are conducted with some rationality and in freedom. As it is, controversialists cannot help employ persuasive means—in particular rhetorical tricks and showmanship. As for freedom of inquiry, which includes the freedom to criticize and publish criticism, as well as to hold unorthodox views, it varies across research fields and from society to society, but it is never complete. Some groups, even within the scientific and humanistic communities, wield more power than others, and so are better equipped to impose their own views. In science most of the time there are illustrated despotisms, but occasionally they stifle legitimate debate. For example, physicists are nowadays less tolerant to unorthodox views than social scientists.

Scientific and technological controversies eventually come to an end if they are purely technical, i.e. if they do not involve any philosophical or ideological principles. They come to an end in any of several ways. Here are some: (*a*) one of the parties exhibits a decisive observation, calculation, or technique; (*b*) a synthesis of the two rival views is proposed, that incorporates the good points of each; (*c*) a third party enters the controversy and wins the day; (*d*) the issue is forgotten for lack of interest; (*e*) censorship stops the discussion and imposes the official view: (*f*) one of the parties brandishes a bunch of computer printouts and announces 'The computer says so'.

There should be no doubt as to the efficiency of criticism to weed out error and thus clear the way for new ideas and even novel approaches. Thus Galilei's polemical works were no less important, for the revolution which he spurred, than his scientific ones. However, we should not follow Popper and overrate the power of criticism to advance knowledge, for the thrust of scientific and technological research is invention and discovery, not argument. In these fields controversy is the salt not the roast.

Besides, scientific or technological controversy, if overdone, can have negative effects. One is that it obfuscates the parties involved, which makes them lower their standards of rigor and objectivity, as well as harden whatever dogmatism may infect them. Controversialists are finally won by the desire to win the argument rather than to find the truth. Another negative effect of controversies is that they often contribute to the fragmentation of a research field. This happens whenever each party sees only one of the sides of a question, with its legitimate problem system, but fails to see the other. One example is the century-old debate, mostly mute, between the biologists who seek laws and those who try to establish life histories or phylogenetic lines. Far from being competing research lines, these are mutually complementary and should be merged instead of being kept separate.

The philosopher can participate in scientific or technological controversies either by taking sides or as an arbitrator—particularly if he feels that the contending parties are talking through one another and should cooperate rather than fight. And the historian's task is neither to hush up controversy for fear that it will spoil his story of cumulative induction, nor—*pace* Agassi's brilliant allegation (1963)—to read the history of science and technology in the same way as that of ideology or philosophy, i.e. as a story of schools and their controversies. In science and technology controversy is healthy provided it is conducted honestly and in freedom, but it is seldom the main occupation of researchers. The main task of scientists is to unveil the world, that of technologists to design means to control it.

4. CONCLUDING REMARKS

One way of gaining knowledge is by making observations or measurements, another is by forming and checking conjectures, and a third is by ferreting it out of available knowledge, i.e. by inference. All inference is from some nonempty set of premises—data, hypotheses, or definitions. True, the set of logical truths can be said to follow from nothing, i.e. to be entailed by the empty set of premises—provided rules of inference are granted. *Ex nihilo nihil fit.*

Inference yields new propositions but not new concepts. Thus if we know the sum and the difference of two numbers, we can find out what those numbers are—provided we have already the notions of number, addition, subtraction, and division. Therefore it is mistaken to call quarks, genes and mental processes 'inferred entities': they are hypothesized not inferred.

We need both rigorous (valid, deductive) and nonrigorous (invalid, nondeductive) inference in all fields of research. Even the most austere research, if fruitful, will beget logically illegitimate brain children—new concepts not defined in terms of the known ones or new propositions that do not follow from the known ones. No one type of inference is more productive than any other type. New concepts are not the product of inference, and hypotheses obtained by rigorous deduction (e.g. in theoretical physics) need not be truer or deeper than conjectures born of sheer insight.

Inference has been regarded as a surrogate of experiment (Rignano, 1923). To be sure some reasoning is thought-experiment and it may save us or, on the contrary, suggest real experiment. (Think of planning chess moves or of playing around with mental images or scale models of molecules, buildings or armies.) But no reasoning can replace experiment: there is no substitute for empirical data. And no experiment can replace reasoning: every experiment is designed and interpreted rationally. Reasoning and experiment are therefore mutually complementary, not interchangeable.

A hasty induction has led empiricist philosophers to proclaim the *inductivist* thesis that every generalization, or at any rate every true (or warranted or believable) generalization is inductive: that we start every inquiry into matters of fact by studying individual cases, and conclude it by summarizing our findings into inductive syntheses, never proceeding from hypothesis to observation. By contrast rationalists, seduced by the success of pure mathematics, uphold the *deductivist* thesis that induction is unimportant or incidental: that we start every inquiry with some hypothesis or other, and find out by deduction what it commits us to; experience would come, if at all, at the very end, and its power would be negative rather than positive: it refutes but does not verify.

We endorse neither inductivism nor deductivism but do adopt elements of each. In our view research is sparked off by problems, not by observation or by hypothesis. Observation poses problems or checks hypotheses, but it is neither the only source of problems nor the only way of testing hypotheses. All inquiry except the most trivial involves deduction mingled with plausible inference, and the latter includes induction but is not restricted to it. Confirmation and refutation are seldom conclusive—except, again, in trivial matters. And induction supplies only relatively modest hypotheses and no theories at all. However, induction is involved in the checking of every generalization. The empirical worth of a hypothesis or theory is gauged not only by its resistance to refutation attempts but also by its positive confirmation or inductive weight (the set of data it accounts for) as well as by its theoretical support (the set of theories with which it jibes). But all this will be discussed at length in Ch. 12, after having studied the notions of problem, hypothesis, theory, explanation, forecast, and their kin.

PART III

EXPLORING AND THEORIZING

CHAPTER 7

EXPLORING

From an epistemological viewpoint, to live is to face and solve problems. We face practical problems, such as that of finding the next meal, and cognitive problems, such as that of finding out the nutritive value of a given food. We tackle some cognitive problems for pleasure and others because their solution may help us solve practical problems. Likewise we handle some practical problems for their own sake and others because their solution may allow us to pose or solve cognitive problems. All higher vertebrates handle cognitive problems, and man excels at inventing new problems on top of those he meets in daily life. And yet traditional epistemology has paid little if any attention to problems, the starting point and object of every inquiry. As a result we know very little about the biology, psychology, and logic of finding and solving problems.

In this chapter we shall peek at exploratory behavior and shall identify some problem sources. We shall also study the ways cognitive problems are selected for investigation. In the process we shall meet two concepts that in a way are mutually complementary: those of intuition and method. This will take us to the question of approaching a problem, which is part of every research. Finally we shall unveil the form and content of problems. None of this will transform the art of formulating problems into a science, but it may advance the art and help us understand the most fascinating of all cognitive operations, namely research.

1. EXPLORATORY BEHAVIOR

1.1. *Psychobiology of Problems*

Many animals seem endowed with an innate exploring or investigating ability (Pavlov, 1927, p. 12). This is the capacity to recognize and tackle problems of various sorts—e.g. "What is that?", "Is that edible?", "Is that friend or foe?", "Where may I find water?", "Can I negotiate that obstacle?", and the like. What is peculiar to man is not the capacity to recognize and tackle problems but that of formulating them consciously and explicitly, and investigating them methodically rather than by blind trial and error. Also peculiar to man, at least over the past 10,000 years or so, is interest in purely conceptual problems removed from practical concerns, and pleasure in questioning solutions as well as in conceiving new problems.

A number of biologists seem to go farther and claim that all surviving organisms have "solved" successfully the "problems" posed by their environments, and that such "solutions" (which others call 'adaptations') constitute genuine inventions. Some even say that biological evolution is a problem solving process. And a few have gone as far as to hold that all a scientist or an engineer can do is to imitate, adapt or improve such "inventions". However, most biologists, if pressed, will adopt a sober attitude and own that all that is metaphor: that, although all organisms face challenges, not all challenges are or pose problems. And most psychologists will agree that only some animals, namely those capable of learning (and therefore knowing something), handle problems proper.

An animal may be said to *face a problem* when it feels the need or wish for some additional piece of knowledge—which of course presupposes that it did have some knowledge to begin with. The problem is *cognitive* if its solution might satisfy a purely cognitive need or wish. It is *practical* if its solution might be employed as a means for attaining a practical goal. In particular, a practical problem is *moral* if its solution might guide behavior that could affect the welfare of some conspecifics.

Like any other distinctions, these should not be transformed into detachments. Thus finding out that a certain drug has serious toxic side effects poses the cognitive problem of finding antidotes to it, or the moral problem of whether it should be banned. And appreciating the miseries of underdevelopment poses the cognitive problem of how best to contribute to alleviating them, as well as the practical problem of implementing whatever measures have been designed to solve those practical problems. In sum, sometimes sheer curiosity or wonder, and at other times need or want, sparks off inquiry, which in turn may result in discoveries or inventions capable of satisfying either. There is then no incompatibility between the culturalist and the economicist hypotheses concerning the root of inquiry: they are mutually complementary rather than exclusive.

An animal engaged in solving a problem explores: it explores the environment or itself, in particular its own store of knowledge and skills. However, not all exploratory behavior is problem solving: only *purposeful exploration* is. Thus the automatic and random exploration conducted by animals incapable of knowing anything is not of the problem solving kind:

not knowing anything, they have no knowledge gaps to fill. And it is doubtful that we are justified in assigning inquiring automata the ability to solve problems, for they have no purpose or design of their own: any intention they may have has been built into them by their manufacturers. They are not really inquiring machines but machine aids to human inquiry. As Duncker (1945) put it, "A problem arises when a living creature has a goal and does not know how this goal is to be reached".

Purposeful exploration, be it of the environment or of oneself, is triggered by the need or wish to cope with novelty, experience new situations, perceptions or feelings, or conceive new ideas. Neuropathology suggests that the arousal of curiosity and the drive to inquire are states or events in certain subsystems of the brain, in particular the mammillary bodies, the limbic system, and the hippocampus. In fact when any of these systems is damaged by accident or by drugs, in particular alcohol and marihuana, the questioning ability is seriously impaired. This impairment is extreme and irreversible in Korsakoff psychosis, characterized by lack of curiosity and initiative: the patient is apathetic, lacks initiative, and places himself at the mercy of his environment. He is good only at repetitive tasks: he cannot face new problems, let alone seek them.

A Korsakoff patient does not cope with novelty but at least he can be aware of it. In deep sleep or coma we may not even perceive, let alone seek, new stimuli: we are not alert. The alertness center of the mammalian central nervous system is the brainstem reticular formation (Moruzzi and Magoun, 1949). This system receives inputs from all sensors, and acts in turn on both the cerebral cortex and the endocrine system, in particular the pituitary-adrenal system. In turn, there is a feedback of the latter systems on the reticular formation, so that arousal is really a neuroendocrine process (Hennessy and Levine, 1979). Moreover the action of the reticular formation is nonspecific: it does not "say" 'Look at that' or 'Smell this', but just 'Watch out: something is up'. It prepares the higher centers of the brain to process the output of the sensors, or keeps them generating new information.

Exploratory behavior may be elicited by environmental events or by internal events, particularly in the brain. In either case exploratory behavior is driven by curiosity, which may be regarded as the drive generated by a knowledge deficit or imbalance. Thus a problem may be construed as the difference between what is known and what one wishes or needs to know.

Animals do not always wait for stimulation to trigger exploratory behavior: all of the higher vertebrates actively seek stimuli of certain kinds, so that stimulation comes sometimes as a result of behavior, not the other way round. (They seek excitation up to a certain point, avoiding it only when it becomes excessive. Roughly speaking, the value of the overall external stimulation is bell shaped, but the maximum shifts with the internal state: to lower excitation values when the organism is tired or sick, to higher excitation values when it is healthy. (See Leuba, 1955.))

The schema of active exploratory behavior is not *stimulus–new internal state–response* but rather *curiosity–exploratory behavior–stimulus–new internal state-response* (Berlyne, 1954). Moreover, as exploration proceeds curiosity may be satisfied—i.e. the epistemic drive reduced. Or, on the contrary, the discovery or invention of new items may augment curiosity and thus sustain exploration or redirect it. Thus, whereas most other behavior types tend to reduce the corresponding drive, exploration may increase it. That is, inquiry is often self-sustained: it keeps generating problems. We call this type of inquiry *research*, to distinguish it from mere *fact-finding*, a process that reduces the epistemic drive as the corresponding problem is solved.

Looking for an address and computing a value of a function are instances of fact-finding inquiry. On the other hand conceiving of a model of the problem-solving process, and designing an experiment to test it, are examples of research. A fact-finding process comes to an end when a solution is found or shown to be impossible with the means at hand. In contrast, research generates new problems as it proceeds: in principle it is open-ended either because the solutions are imperfect or because they allow one to pose further problems. Most computer programs are of the fact finding type and a few of the research kind, but none is designed to pose new problems. Computers are problem solving aids, not problem finders.

A problem is a knowledge gap, and a problem solving process is one aiming at filling such gap: it has a very definite purpose. If we knew more about the neurophysiology of questioning and learning we would be able to picture this process with some precision, as the trajectory of a point in a cognitive space, from formulation to solution. See Figure 7.1. The starting point is included in the region of the cognitive space constituted by all the items relevant to the formulation of the problem—i.e. what may be called the *problem setting* or *framework*. Unfortunately, we do not even know which variables to take as axes for building such cognitive spaces.

Presumably the cognitive space of a newborn is very small, exclusively sensory-motor, and nearly amorphous. As the child grows up and learns,

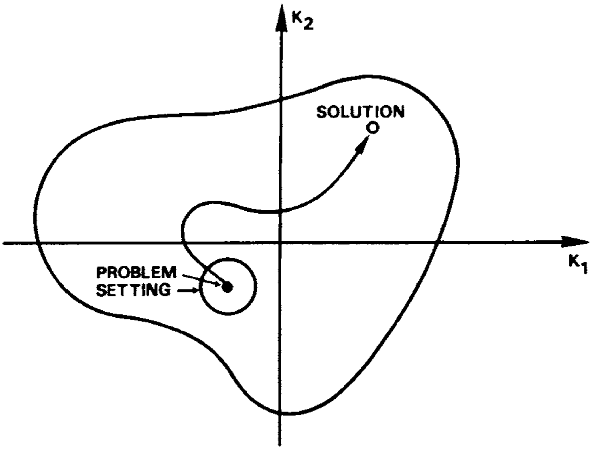


Fig. 7.1. Cognitive state space, and problem solving as trajectory of point representing state of knowledge, from formulation to solution. The axes K_1 and K_2 are (ranges of) functions representing neurophysiological variables characterizing the particular neural system performing the cognitive task in question. The cognitive states in the neighborhood of the problem constitute the problem setting, i.e. the collection of knowledge items relevant to the formulation of the problem.

its cognitive space expands qualitatively (new axes) as well as quantitatively (wider regions), and it becomes structured—i.e. the various knowledge items get interrelated forming systems. Such systems act as frameworks or settings for new problems. Thus when confronted with a new problem we are not always at a loss, but can often place the problem in our cognitive space and reconnoitre its neighborhood (setting) in search for clues. To change the metaphor, the inquirer forms, stores and organizes "frames" or data structures representing stereotyped situations, like moving about in a school building; such frames and frame systems are ready to receive or generate new experiences (Minsky, 1977).

All hard problems demand dedication, often over very long periods. However, it is common experience that continued exclusive concentration on a single problem does not pay: one is likely to fall into a rut that may lead to a *cul de sac*. Distractions are useful at this point: either turning temporarily to lighter problems, or just going for a walk. This manoeuvre has sometimes been interpreted as shoving the problem under consciousness and letting the Unconscious do the work. However, the recourse to the inscrutable Unconscious is not only a refuge in ignorance but also quite unnecessary, for there is a better explanation.

Concentration on a single problem tends to strengthen certain neural pathways while inhibiting others. (Recall Ch. 1, Section 1.1.) Distraction loosens some connections and gives others a chance to be established. This allows us to get off the trodden path and explore alternatives. When not under the pressure to move along the rut one may wander more freely in the cognitive space, which may eventually lead to the desired solution. Such happy end would not occur if we had no definite problem in hand (or rather brain) and some competence to handle (or rather mind) it. Neither the Unconscious nor blind chance (pervasive though it is), but just the leisurely exploration of new possibilities, does the trick.

In general, repeated experiences facilitate the formation of routines and *eo ipso* impair the ability to solve and even "see" problems of new kinds. Again, the use-disuse hypothesis explains this phenomenon: the neuronal connections become stronger, the more they are exercised. Hence experience, when confined to a few kinds, tends to decrease plasticity, whereas it tends to enhance it when varied. When a routine develops the tendency to disregard problems of new kinds, or to "assimilate" them to those of old kinds, it becomes an obstacle. One tries his old tricks on the new problems and, more often than not, one fails. This phenomenon of crystallization is what the Gestalt psychologists used to call *Einstellung* (problem solving *set*), and Duncker (1945) *functional fixedness*. (Cf. Katona, 1939; Birch and Rabinowitz, 1951.) People with fixed tasks become more and more rigid in their ways, until they become capable only of reproductive not productive thinking (Wertheimer, 1961). *Moral*: When scouting for someone capable of facing a tough new problem do not advertise for an experienced specialist but for a newcomer willing to take the risk and jump over disciplinary frontiers. He may undergo the neural reorganization we experience as the flash of insight (or *aha* experience).

By themselves, then, learning and praxis do not improve questioning ability or problem solving competence. The effect of both depends largely on the inquirer's attitude and opportunities. If his goal is to maximize his knowledge store or his practical utility, he will absorb solutions more readily than generate problems. On the other hand, if his goal is to maximize his contribution to knowledge, he will prefer unsolved problems, whether ambitious or modest. However, even the research-oriented inquirer—the one who hops from problem to problem instead of accumulating solutions—faces at all times the danger of becoming an erudite. Indeed, as science, technology and the humanities advance, it is only natural, as well as reasonable, to resort to the stock of solved problems when first confronted with a new problem. This tactic is bound to succeed if the new problem is indeed similar to an older solved problem. But if the problem bears only a superficial analogy to an older solved problem, the erudite's procedure will lead nowhere or, worse, it will yield a wrong solution, or will generate more useless problems than it solves. (Adams (1980) calls this the 'hit-and-run' approach.)

The questioning ability, the capacity to "detect" new problems, can be trained or inhibited. The training can be of either the argumentative or the investigative type. The former consists in encouraging questioning and arguing without engaging in any inquiry beyond some fact finding—usually the spotting of the proper chapter and verse. This is the training that theology students, and even many humanities students, receive in the hope of discovering the truth by exegesis of canonical texts and by argument. As we have seen (Ch. 6, Section 3.1) it serves the entrenchment of dogma rather than the generation of new knowledge. The best training of the problem finding and solving capacity is one in mathematics, science, or technology. Indeed in these fields original work consists in investigating open problems with the aim of coming up with new results. Not that criticism is absent from such training, but it is only one among several means for producing or at least controlling new knowledge.

The questioning ability can be inhibited by repeatedly declining to answer the child's natural questioning, by giving unintelligible or false answers, or by imposing the uncritical acceptance of what is regarded as the right view. Unfortunately too many schools all over the world discourage questioning: they are *dressage* schools aiming at the training of conformists. (As cosmologist Hermann Bondi has stated, the few failures to this training in question-stopping are called 'scientists'.) There are three obvious reasons for this kind of schooling. One is that a questioning child is a source of additional work or even trouble. Another is the belief that all a child must learn are "facts" (data) and skills—questioning not being among the latter. A third is that questioning can be worse than annoying: it can be subversive, and most of the people who wield power do not want society to change (Ackoff, 1978, pp. 4–5). As a South American minidictator said, there is but one step from the scientific finding that matter is unstable, to the suspicion that even the most sacred institutions could be unstable as well.

Because of the obsession for results or findings, the problem of the psychological mechanism of the emergence of problems has hardly been studied. All we seem to know is that inquisitiveness is innate in many species and that it has a positive survival value. On the other hand there are a few hypotheses about the problem solving process. The oldest is perhaps that problem solving proceeds by association. Although this view may explain (in a superficial way) why some ideas are recalled at the proper time, "it cannot explain the appearance of other ideas, e.g., original ideas" (Meier, 1931).

Most of the workers in this field are either empiricists or cognitivists. The former hold that, when tackling a problem, the animal proceeds by trial and error and gropes gradually towards the solution, i.e. reaches it through decreasing errors. This is at best a description of problem solving behavior. It does not explain the process because it does not tell us how the tentative solutions (hypotheses) are formed or why the process should converge. (A genuinely haphazard trial and error process should hit on the solution only by accident: there should be no gradual diminishing of discrepancies.) On the other hand cognitivism claims that the animal starts out its exploration with a whole set of hypotheses in mind, or else hits on the correct one in a flash of insight.

There may be something in each of the above rival views, and undoubtedly both of them miss something. The animal's behavior in the face of a new problem depends on its experience in handling problems of a similar kind, as well as on its goals and expectations. If it has no experience whatever, it will either avoid the problem altogether or tend to proceed by trial and error until reaching a satisfactory solution or giving up in fatigue or boredom. If on the other hand the animal does have some experience with similar problems, it will start out with a fund of guesses and try them out in succession. In either case, though, the subject keeps forming and testing new guesses. Moreover the process is not totally haphazard because there is feedback from some outcomes. I.e., unless the proposed solutions are totally irrelevant, or their adequacy kept from the subject, he may learn from his mistakes—not only learn to give up some of the wrong answers but also in what direction he should advance toward the correct response.

There is no finding and solving of problems without some motivation, but there are many kinds of motivation. Even in the case of purely cognitive problems, such as those in mathematics, basic science, and the humanities, the individual researcher may be driven not only by curiosity but also by love or hatred, celebrity or power, security or money. (My own motivation in writing my best physics paper was political.) So much so that, when any such noncognitive goal is reached, the individual may no longer feel motivated to pursue his research. (Most research careers terminate in a doctorate, a professorship, or a deanship.) Yet no extracognitive drive, however strong, will accomplish anything of cognitive value unless accompanied by strong curiosity helped by talent, imagination, and industry.

Nor is a strong epistemic drive, allied to competence, sufficient to succeed in problem solving activities. Three further factors are involved: aiming at certain collective goals, choosing the right problem, and luck in investigating the problem at the right time in the right inquiring community. The collective goals are those of the given field of inquiry: they are not impersonal but are in the intersection of all the sets of individual motivations of the workers in the given field. Thus the collective goal of mathematics and basic science is the *finding of patterns*; that of technology, the *designing of useful artifacts*; that of the humanities, the *understanding of culture*; that of the liberal professions, *serving needs* of others. No matter how selfish a worker in any of these fields may be, he must aim at one of those collective goals if he is to make any contribution to it. So much for the collective goals of inquiry. As for choosing the right problem at the right time in the right circumstances, this matter deserves a separate section.

1.2. *Sources and Choices*

The traditionalist approach to learning presupposes that basic (or perhaps all of) human knowledge is locked in certain texts, whether sacred or secular. This is the approach adopted by religionists, schoolmen, hermeneutic philosophers, and doctrinaire ideologues. It follows that learning consists in studying and suitably "interpreting" such texts, and trying to solve every new problem by "applying" the suitable passages. (Failure in solving problems is blamed on the "interpreter" not on the texts.) In short, in this approach there can be no genuine new knowledge, and this because no new problems are admitted.

Modern workers in mathematics, basic science, applied science, technology, and the advanced sectors of the humanities, make no such presupposition and therefore proceed in a radically different manner. They presuppose, on the contrary, that nobody knows everything, that much of what little we know becomes quickly obsolescent, that our natural and social environments keep posing new challenges to be conceptualized as problems to be investigated, and that certain inquiries consist in posing and solving problems with no immediate practical roots or implications. In short, research-oriented inquirers know that *inquiry consists in investigat-*

ing problems. And they know that there are two problem sources: *external*, or originating in the natural or social environment, and *internal*, or originating in our own fund of knowledge—or rather in its gaps.

Talk of problem sources should be taken metaphorically not literally, for problems do not spring before us the way water springs from an artesian well. We do not discover problems the way we discover iron ore deposits: problems, like their solutions and, in general, all conceptual objects, are *made* by us. To be sure, not all environmental challenges arise spontaneously: some result from mismanagement of natural or social resources. However, man has lived for many millennia without transforming such challenges into those peculiar neural processes we call 'problems'. External challenges are occasions or opportunities, not causes of inquiry. Unless there are well equipped, trained and motivated brains, no challenge, however pressing, will elicit the conception of the corresponding problem, let alone of a research project to investigate it. Thus even now most people do not realize the gravity and urgency of the challenges to the very survival of the human species—starting with the arms race—so they keep postponing facing them, i.e. transforming them into problems to be investigated and solved.

So, it is simply not true that the environment, in particular society, exudes problems: individual brains, immersed in a suitable culture, create them. Nor is it true that every single problem we do manage to pose and solve is nothing but a response to challenges posed by "the historical development of productive forces and production relationships", as an early advocate of sociologism put it (Hessen, 1931, p. 203). Most problems in logic, pure mathematics, physics, astronomy, anthropology, history, and philosophy—to mention but a few fields of inquiry—are internal, not external. Workers in these fields handle mostly problems originating in what used to be called 'idle curiosity'—a sin to be shunned by every properly educated child. Let us praise and promote idle curiosity, for it prompts the most noble and, in the end, also the most useful of activities: inquiry. As Nobel laureate George Wald said, "The great questions are those an intelligent child asks and, getting no answers, stops asking. That is known as growing up." (Quoted by Teuber, 1966, p. 583).

Obviously, there is an enormous variety of internal problems as well as a number of kinds of stimuli and inhibitors to the problem finding activity. See Table 7.1.

Having emphasized that many problems in basic research originate in disinterested (idle) curiosity, let me hasten to acknowledge that others are

TABLE 7.1
Some stimulants and inhibitors of problem finding and solving

Stimulants	*Inhibitors*
Exploring new territory	Censorship (by schools or ideologies)
New facts to be checked or explained	Routine
Anomalies or contradictions	Complacency
Moderate criticism	Destructive criticism or indifference
Moderate competition	Rat race
Group cooperation	Infighting
Social rewards	Absence of social rewards

offshoots of practical problems, or have solutions that can be used to solve practical problems, whereas still others are elicited by environmental situations. An example of the first kind is the wealth of mathematical and scientific problems posed by technological developments; thus hydrodynamics owes more to aeronautics than the other way round. An example of the second kind is the utilization of biology in medical research and practice. An example of the third kind is social science: here the choice of problems is largely motivated by contemporary events. Let us take a look at this point.

Coleman (1980) has shown how the Chicago school of sociology, which flourished in the first quarter of this century around Robert Park, was mainly interested in the problems associated with the urbanization of North America (slums, ghettos, gangs, prostitution, etc.). Likewise the Columbia school built around Paul Lazarsfeld tackled some of the problems that arose in the 1930s, mainly the formation of a national market and of a national communication network. These problems called for a methodological shift, from the study of small groups (e.g. urban subcultures), to samples of large regions, sometimes national samples. Finally, the third or contemporary phase appeared with the beginning of social policies (as distinct from economic policies) in health, education, welfare, employment, etc., in the late 1960s. This research into social policies involved process evaluation, product evaluation, and even some large scale social experiments. It is definitely a shift towards applied social science. (Interestingly enough its leader, Professor Coleman, started out as a mathematical sociologist.)

So, there is two-way traffic between the internal and the external sources of problems. Besides, like every other human endeavor, problem finding and solving are subject to chance. Sometimes chance events spark off problems or solutions, at other times they nip in the bud problem solving processes. A book on probability, seen by chance by a meteorologist, gives him the idea of investigating random processes in the atmosphere. A film on alcoholics gives a neurophysiologist the idea of pursuing a certain clue regarding taste formation and aversion. A chance encounter between a mathematician and a psychologist suggests to the former a new problem and to the latter a solution to his current problem. Computers are protected against such chance encounters: they are always programmed. So, they are neither led to truth nor misled into error in a haphazard way. Human brains, on the other hand, are biased cognitive roulettes.

The bias lies usually in the curiosity of the researcher or in the urgency of the problem. But sometimes fashion is involved; some problems are fashionable even when unimportant, whereas others are unfashionable even when important. One cause is obvious though unjustifiable: most middle class people like to be thought of as gentlemen or ladies of fashion. (Nowadays it is unfashionable to say that Bach and Wittgenstein are boring, that psychoanalysis and much of cosmology are unscientific, or that the abuse of computers is a threat to science.) Another cause is less obvious and thoroughly justified: science and technology must be up to date and, justifiably, this draws many people to work on recently posed problems.

Finally there is the question of timing. The alchemists posed too soon the question of the transmutation of metals, and the monetarists are too late in advocating a return to the competitive (free) market as the solution to the problems of the advanced capitalist economies. As Nobel laureate Sir John Kendrew said, "One has to ask the right question at the right time". But what is a right question? And can we say beforehand whether a question is good or bad, or must we wait and see what impact research on it has had? Let us deal with one question at a time.

Clearly, not all problems are worth the same. Thus finding out the neural mechanism of problem formation has a far greater cognitive importance than discovering a huge new deposit of mineral oil, but the practical importance of the latter may be far greater than that of the former. So, when ranking problems (or anything else) we should start by specifying the respect in which we are doing it: cognitive or practical. And we should keep in mind that such categorization may be only temporary, for some problems that start out being "academic" may turn out to be practically important (such as the problem of nuclear forces), whereas some practical problems (such as whether or not trenches for warfare should be straight) may disappear before being solved.

Once this has been done, how shall we assess the relative importance of problems in the given respect? Will difficulty be a suitable measure, or perhaps the length of time that the problem has been around? The former is not a good measure, for difficulty is largely subjective: some good mathematicians experience considerable difficulty in performing elementary sums, whereas computers can perform calculations beyond the reach of any brain. Nor is length of time—particularly when measured by the number of pipes smoked by Sherlock Holmes—a good indicator of problem importance. Thus the four-color problem was solved only recently after two centuries of unsuccessful trials, but it is neither theoretically nor practically important: it is a stray problem and presumably its solution will not elicit a spate of other mathematical problems. We must look for alternative criteria.

We may start by classing problems into good and bad according to this definition: A problem is a *good* problem if (*a*) its solution would make some contribution to knowledge or human welfare and (*b*) it can be solved with existing or conceivable means. But of course not all good problems are important: some are trivial, others mediocre, and a few exceptional. We stipulate that a good problem is *important* if (*a*) it belongs to a problem system (rather than being stray), (*b*) it is deep, i.e. presuppose or is likely to upset basic principles or traditional behaviour patterns, and (*c*) its solution is likely to make it possible for a number of other people to pose further good problems.

By using the above definitions we can estimate a priori the value of problems. However, here as elsewhere our estimates may be erroneous. In any case three things are clear. One is that all research proposals are evaluated both a priori and a posteriori, and prior evaluation rests on some more or less tacit idea of problem worth. Second, if these ideas of problem worth were rendered explicit and discussed in the open, project evaluation would be more rational and objective, and therefore more fair. Third, very few researchers have a knack for "spotting" important problems, even fewer the ability and drive needed to solve them. Most researchers are better at problem solving than at problem having. (Worse, many researchers do not even seem to know that research consists in working on problems—so much so that, when writing their final reports, they seldom start by stating their problem.) Only exceptional investigators, such as Newton and Hilbert, are capable of drawing lists of important problems nourishing several generations of workers. This requires flair as well as experience. (Cf. Bunge, 1962.)

Estimating the worth of problems with hindsight is easier than doing it a priori, but there are pitfalls here too. First, historians know that it is easy, perhaps unavoidable, to write "whig history"—i.e. to view the past against the background of our own problems and aspirations, and so to overrate certain accomplishments of our predecessors (mainly those which we have used) and underrate others. Second, even an objective indicator of worth, such as the volume of publications devoted to a problem, is not completely reliable, for there are always extrinsic factors, such as fashion, ideology, and availability of resources. Moreover in certain fields, such as philosophy, the relation between the importance of a problem and the volume of the publications devoted to it is inverse, not direct. Thus in recent philosophy trivial problems, such as the status of counterfactuals, the raven paradox, and the way to identify individuals across (logically) possible worlds, have attracted far more attention than the concepts of problem, partial truth, or attribute.

Although there are no rules for generating good problems, it is possible to give some helpful advice in this matter. First, *reexamine old problems*, for some of them may have changed just because their setting has altered, so that they can now be restated and solved in a different way. For example, the entire problem system of genetics was altered by the emergence of molecular biology and, in particular, the discovery that DNA is the genetic material. Also, practical problems "do not stay solved" (Ackoff, 1974) because society changes. Thus the automobile solved the urban transportation problem at the beginning of our century but is no longer practical.

Second, *examine critically* the known solutions to old problems, for there must be some flaws in some of them. (Enrique Gaviola, personal communication.) For example, by repeating a run of measurements or calculations, or reanalyzing a set of assumptions, one is bound to find something to correct or improve, generalize or discard. Thus we still do not have a satisfactory derivation of thermodynamics from statistical mechanics, or a good theory of liquids.

Third, *apply known methods or theories* to new problems, and check whether they still hold. If they do, you will have extended the domain of those methods or theories, and if they do not you may have generated a whole new problem system. For example, the Newtonian theory of gravitation, originally devised to explain the solar system, was successfully extended beyond it—until its own limits were discovered.

Fourth, *generalize* old problems: try introducing further variables or more complicated relationships among them. For example, introduce nonlinearities or inhomogeneities in equations, or impurities or external perturbations in concrete systems.

Fifth, *dig deeper*: try to obtain known results from assumptions concerning deeper levels. For example, try to explain psychological laws in neurophysiological terms, or historical trends in terms of social and economic laws.

Sixth, *look outside*: search for relationships to problems belonging to different fields. For example, explain chemical reactions as nonelastic scatterings of the reactants, or relate economic activities to political and cultural ones.

Seventh, *go out of fashion*: choose tough and interesting open problems that are not being currently investigated by others.

If all of these recipes fail, keep searching the literature, consult with others, ask someone for a problem, or take a holiday. Should all these moves fail after a couple of attempts, give up: this is far better than simulating research when actually doing nothing but keeping busy with books, instruments or computers.

An experienced and productive investigator needs no advice on how to generate new problems, for he stumbles on them all the time, and he may be in a position to give such advice. Problems seem to arise spontaneously in the course of his research, and to do so beyond his control and often to his own surprise. Sometimes the new problems are so numerous that he has to farm them out to assistants or students. For example, in measuring a magnitude a scientist may discover a fault in the experimental design or its implementation, which flaw generates the problem of overcoming it. In the course of the investigation of this problem he may hit on a new effect or a new technique, and so he may get sidetracked: the derivative problem may prove to be more pressing or even more interesting than the original one. Or in building a mathematical model of some thing or process, a scientist or technologist may formulate a new equation posing a new mathematical problem; conversely, in studying some mathematical work the idea may occur to him of trying the newly learned mathematics on one of his substantive problems that had proved to be intractable before. In sum, original research, unlike routine research, is an *open ended and self-sustained process.* If and when it does grind to a halt it is because people have lost their curiosity or the freedom to satisfy it.

2. FROM INTUITION TO METHOD

2.1. *Intuition*

When faced with a problem of a familiar kind we resort to our own fund of knowledge, or to the knowledge stored in other people's brains or in libraries. This will not suffice if the problem is of a new kind: in this case we need additional knowledge and some intuition or flair to guide us in the search for such knowledge. Intuition or insight is of course that proteic and ill defined ability to spot problems or errors, to "perceive" relations or similarities, to form concepts or hypotheses, to conceive of strategies or tactics, to design experiments or artifacts—in short to imagine, conceive, reason or act in novel ways. (Cf. Bunge, 1962.)

The novelty occurring in a flash of insight is of either of the following kinds: *radical* (unprecedented) or *combinatorial* (the result of combining previously known items). Radical novelty in ideation consists presumably in the formation of a psychon (plastic neuronal system) of a new kind, whether spontaneously or under the influence of the activity of some other psychons. On the other hand combinatorial novelty in ideation, i.e. association, consists presumably in the linking of two or more activities in so many neuronal assemblies (Hebb, 1949, p. 134). As long as the latter remain isolated from one another, the two ideas stay apart. But if, perhaps by chance, a bridge is formed between them, then the relation between the two formerly separate ideas emerges suddenly as a third idea. In both cases chance, as well as the deliberate scanning of a region of the cognitive space, are likely to occur.

The newborn has no insight: he must build it up. However, sheer quantity and variety of experience are not sufficient to acquire a powerful intellectual intuition although they are necessary for it (Birch, 1945). Hard work of the right kind, success, and luck—i.e. suitable circumstances—are needed as well. Hard work of the right kind—tough tasks with scant means—is necessary for it cannot be performed in a routine fashion but calls for imagination. Success is needed because repeated failure discourages and counsels turning to easier tasks. And luck is needed because adverse circumstances beyond our control may prevent us from honing our wits. Thus too long hours applying rules of some kind, or making routine calculations or measurements, will blunt one's intuition. On the other hand repeated exposure to varied problems requiring more ingenuity than specialized knowledge will hone it. For this reason Dreistadt (1969)

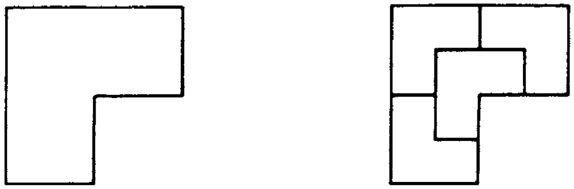


Fig. 7.2. The Farm Problem: divide the L shaped lot into 4 parts with the same shape and size.

recommends proposing problems such as the Farm Problem to cultivate problem solving skills. See Figure 7.2.

No problem can be understood and handled successfully without some intuition or "vision" of it as a whole and against its setting. This holds for all problems, whether cognitive or practical, in every field of human endeavor, from science to technology, from art to the humanities, from management to politics. However, the need for intuition in every field of human endeavor does not entail that is suffices, or that we cannot go beyond it, or that it is superior to experience and reason.

Intuition is insufficient: it is not a substitute for work but a guide to it. There is no substitute for hard work except of course soft work—but this one does not really work. The so-called "aha experience" (or "eureka phenomenon") does not come out of the blue but after hard thinking or doing, after trying different hypotheses or methods, and repeatedly reorganizing one's ideas. (Cf. Gruber, 1981.) Divination is no short cut either: it is no substitute for exploration in the light of specialized knowledge. For example, the rate of success of dowsers (water witches) is far inferior to that of geologists and water engineers: "We know of no other hypotheses that have been put forth so persistently over such a long span of years with such surprisingly negative experimental findings as the hypothesis that water witching 'works'" (Vogt and Hyman, 1959, p. 82; see also Randi, 1979).

Intuition is preliminary and tentative rather than final: it must be elaborated and checked. Moreover mathematics, science, technology and philosophy keep introducing counterintuitive or implausible concepts, theories, and methods. (Actually every radically new idea, even if initiated intuitively by its creator, is likely to sound counterintuitive to all others.) In short, intuition is indispensable but intuitionism, the epistemological doctrine according to which it is sufficient, is false. (For a criticism of philosophical intuitionism see Bunge, 1962.)

Counterintuitive ideas are particularly conspicuous in contemporary science and technology. This is so because appearance is only skin deep, and reality so complex that it can be grasped and controlled only with the help of very sophisticated instruments. Photons are not tiny particles, solid bodies are not continuous, organisms are not machines, and national economies are not households. Actually we need not go far out into modern science and technology to find problems and solutions that defy intuition. Recall the story of the king who promised to reward one of his subjects with a few grains of wheat: one for the first square of the chess board, two for the second, four for the third, eight for the fourth, and so on. He had not realized that the total number of grains promised was $2^{64}-1$, which is more than the annual world wheat crop.

As long as we stay on an intuitive level we are constrained to proceed by guessing, also known as the trial and error "method"—actually the very antithesis of a method. To be sure, this procedure is bound to succeed sooner or later in the case of problems involving a manageable number of alternatives. Thus Edison and his coworkers hit on the right material for manufacturing light bulb filaments only after having tried haphazardly nearly 10,000 different materials—practically all those known in his time. (He could have proceeded methodically if solid state physics had existed in his time.) But the hit-or-miss procedure is unlikely to succeed with complex problems even after repeated trials. Electrodynamics and the dynamo were not outcomes of blind trials but of hard and sustained theoretical and experimental work. The alchemists tried for thousands of years to turn lead and other "base" metals into gold by combining and treating at random a number of substances. Nothing of the sort could possibly have worked: a wholly different approach was needed to solve that problem. We shall elucidate the obscure and important notion of an approach in Section 3.2 after having examined the concept of a method.

2.2. *Method*

We are all familiar with a number of methods, or regular procedures for handling problems of restricted kinds, in all sectors or human endeavor, from mathematics to butchery. For example, neuroscientists use the Golgi staining technique, in conjunction with microscopy, to render individual neurons visible. They also employ microelectrodes to record the electrical activity of single neurons, and electroencephalography to record the activity of entire regions of the brain. All these are *techniques*, or special methods. Every technique has a restricted scope, i.e. it cannot be exported to distant fields of research.

Philosophers have been acquained since Antiquity with three conceptual methods: Socrates' maieutic method, Porphiry's tree questioning, and the method of hypothesis. (We do not rate contemplation as a method, for there are no rules for contemplating in a productive way.) The maieutic method supposedly consists in extracting knowledge, by clever questioning, from the student's inborn storehouse of knowledge. Actually the teacher must know the answer and must know how to supply clues that will allow the student to exploit his own background knowledge. The method works for purely logical problems but does not help solve problems in, say, statistics, biochemistry, or history. Tree questioning consists in surveying the set of possibilities and dividing it step by step into mutually disjoint subsets until the wanted subset is reached. It works whenever the subject is initially given a finite set of possibilities and has the means to eliminate the wrong ones—which is not often the case in original scientific or technological research. Finally, the method of hypothesis, or hypothetico-deductive method, consists in the sequence *invention* (of idea or device)—*trial*—*evaluation*. This method is used in all cognitive fields, including ordinary knowledge. And it is used systematically in mathematics, science, modern technology, and philosophy, where hypothetico-deductive systems (theories) are plentiful. See Figure 7.3.

Not all techniques are alike in accuracy, foundation (or justification), and cost. For example, to detect defects in the hull of a ship, X-rays and ultrasound are more accurate and better founded than banging with a hammer—but they are also more expensive. Again, forecasting the state of the economy by divination or on the basis of some economic ideology is inferior to extrapolating recent trends; and forecasting on the strength of well-confirmed theories and hard data should be superior to trend extrapolation.

A technique or special method may be called *scientific* if (*a*) it is intersubjective in the sense that it gives roughly the same results for all competent users, (*b*) it can be checked or controlled by alternative methods, and (*c*) there are well-confirmed hypotheses or theories that help explain, at least in outline, how it works. A method that complies with only one or two of these conditions may be called *semiscientific*, and one that complies with neither *nonscientific*. Note that, whereas condition (*b*) is that of testability (checkability), clause (*c*) is that of foundation or justification, as opposed to faith or authority.

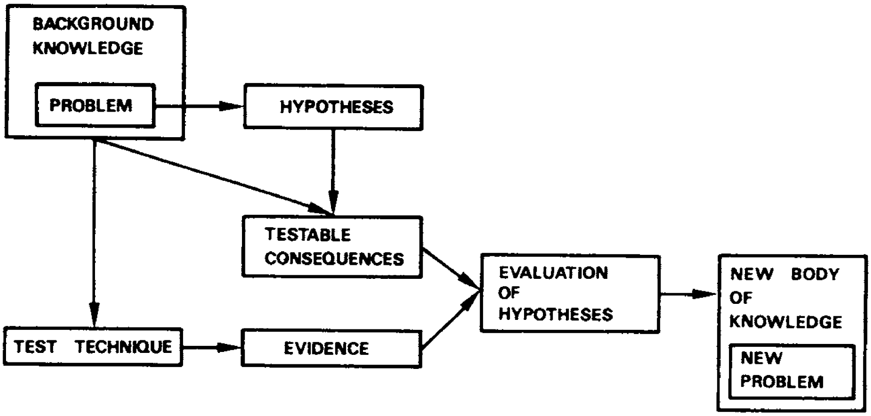


Fig. 7.3. A scientific research cycle. The importance of a scientific research is gauged by the changes it induces in our background knowledge or by the new problems it poses. Adapted from Bunge (1967a).

For example, whereas random sampling is scientific, the ink-blot test is not. Indeed the former is objective, can be checked by investigating the whole population, and is founded on the probability calculus. On the other hand the ink-blot test is subjective and mysterious. A good example of a semiscientific technique is the memorization of nonsense syllables, widely used to study verbal learning. It is objective and can be checked by alternative techniques, but it rests on a false hypothesis. The hypothesis is that, by memorizing nonsense syllables such as *giz* and *toc*, the "blank slate" condition is reproduced, so that learning does not depend on previous knowledge. But this is false: in order to memorize such syllables the subject is likely to associate them with familiar words or images—e.g. 'giz' with 'gizmo' and 'toc' with 'woodpecker'. There is no pure verbal learning. Like all learning, verbal learning is rooted in previous knowledge: it is a growth process, not one of haphazard piling up of unrelated items.

The inception of a new technique may revolutionize a field of human endeavor or it may even give rise to a new field. Think of triangulation, printing, the microscope, chromatography, or computers. The design of a new technique is, increasingly, the work of a sequence or of a team of specialists composed of scientists, technologists, and craftsmen. Much effort is being devoted in science and technology to the invention, improvement, and adaptation of new techniques. However, it should never be forgotten that techniques are means not ends: means for knowing things or getting things done. The goal is either a bit of knowledge or a commodity.

The learning of techniques is an important part of the training of experts. It is also the easiest part, for one need not know why or even how the technique works in order to use it competently. (For example, nobody seems to know how the Golgi staining technique works.) Therefore most specialists are skilled technicians: performers rather than musicians, color mixers and dabblers rather than painters, microscopists rather than cellular biologists, electrode pushers and recorders rather than neuroscientists, calculators rather than mathematicians, and so on.

When on their own, technicians are likely to engage in routine work, haphazard tinkering, or even wasteful activity. Technicians perform best when they are components of teams led by productive artists, technologists, scientists, or managers. This is so because in such case their work services a long-range goal and receives the feedback necessary for correcting false moves. As a component of such a team a technician's tinkering may lead on to a new technique or to the discovery of a previously unknown category of facts.

Some techniques born with a narrow scope have evolved eventually into general methods. Microscopy, the method of successive approximations by iteration, and even the scientific method are cases in point. Indeed microscopy is now being used not only in biology, where it was first applied, but also in physics and various branches of engineering (e.g. metallurgy), as well as in some branches of the humanities (e.g. art history). The method of successive approximations by iteration is employed in all of applied mathematics and is a paragon of scientific procedure. One starts by computing or guessing a first approximation. This approximate solution is then fed back into the original problem to yield a second solution, and so forth. In this way more and more accurate solutions are obtained: the solution is the limit of the sequence. Every partial solution is built upon the preceding ones, and the increase in accuracy or degree of truth is uniform. That is, the iterative procedure is constructive and convergent.

We mentioned the scientific method as one with a large scope, in contrast with any of the special methods or techniques. What is it? Opinion varies from radical skepticism, which denies its existence, to dogmatism, which believes it to be a set of simple, invariable and infallible recipes for finding definitive truths. We take the *scientific method* to be the following ordered sequence of cognitive operations:

1. *Identify a problem* (whether gap or indentation) in some body of

knowledge—if possible an important bit of ignorance. If the problem is not stated clearly, go to the next step, otherwise to the subsequent one.

2. *State the problem clearly*, if possible in mathematical terms or in terms of measurement operations.

3. *Search for information, methods or instruments* likely to be relevant to the problem—e.g. empirical data, theories, methods of measurement or computation, measuring instruments, etc. I.e. scan what is known to see whether it can help solve the problem.

4. *Try to solve the problem with the help of the means collected in the previous step.* Should this attempt fail, go to the next step; if not, to the subsequent one.

5. *Invent new ideas* (hypotheses, theories, or techniques), *produce new empirical data*, or *design new artifacts* that promise to solve the problem.

6. *Obtain a solution* (exact or approximate) of the problem with the help of the available conceptual or material means.

7. *Derive the consequences* of the solution thus obtained. If the solution candidate is a hypothesis or a theory, compute predictions; if new data, examine the effect they may have on existing ideas; if new artifacts, assess their possible uses and misuses.

8. *Check the solution.* If the solution candidate is a hypothesis or a theory, see how its predictions fare; if new data, try to replicate them using alternative means; if new techniques or new artifacts, see how they work in practice. If the outcome is unsatisfactory, go to the next step, otherwise to the subsequent one.

9. *Correct* the defective solution by going over the entire procedure or using alternative assumptions or methods.

10. *Examine the impact* of the solution on the body of antecedent knowledge and state some of the new problems it gives rise to.

Table 7.2 exhibits schematically the scientific treatment of three typical problem types: one theoretical, one experimental, and a third one technological.

The scientific method has some points in common with the method of successive approximations by iteration. First, scientific research proceeds *gradually*: even the correct insights that may be hit upon by chance, and revolutionary new ideas, are the outcome of previous work, as well as subject to correction. Second, scientific research yields, at least with reference to the world of facts, *partial truths* rather than complete and accordingly final truths. Third, the scientific method, in contrast to the hit-or-miss procedure of common sense and unbridled speculation, is *self-*

TABLE 7.2
Three typical problem types.

Step	*Empirical problem: measure*	*Theoretical problem: explain*	*Technological problem: design*
1	What is the value of X?	Why does X have the value x?	How can X be done?
2	What is the measured value of X to within error e?	What premises entail that X is worth x?	How can we design a Y capable of doing X with maximal efficiency?
3	What experimental arrangement(s) Y allow to measure X with error less than e?	What theories Y, subsidiary hypotheses h, and data d imply that the value of X is x?	What information can help design Y?
4	Perform measurement of X with means Y. If Y proves inadequate go over to step 5, otherwise to 6.	Compute the value of X with the help of Y, h and d. If the result is implausible go to step 5, otherwise to 6.	Design Y with the help of the information gathered in previous step. Should design prove inadequate go to 6.
5	Try new technique Y'.	Try new theory Y', or new subsidiary hypotheses h', or new data d'.	Try new design.
6	Use Y' to measure X.	Compute the value of X with Y', etc.	Build an artifact (or a scale model) Y' or run a computer simulation based on the design.
7	What does the outcome of 6 imply or suggest?		
8	Measure X with alternative techniques. If the new result is implausible go to step 9, otherwise to 10.	Compute X with alternative assumptions or methods. If the new result is unsatisfactory go to step 9, otherwise to 10.	Put Y' to some uses and misuses. If Y' does not work satisfactorily go to step 9, otherwise to step 10.
9	Look for systematic errors and correct them.	Look for possible sources of error and correct them.	Look for flaws in design and correct them.
10	How does the new result affect knowledge or practice, and what new problems if any does it pose?		

BACKGROUND
KNOWLEDGE
PROBLEMS
METHOD
SOLUTIONS

Fig. 7.4. A method as an input-output box. Input: problems and background knowledge; output: solutions.

correcting: it can help recognize errors and obtain higher order approximations, i.e. truer answers. One difference between the scientific method and the mathematical iterative procedures is that the former never allows one to find exact solutions of nontrivial problems. Another difference is that the scientific method can be applied to all inquiries, whether mathematical or empirical, scientific, technological, or humanistic. (Some philosophers deny this: we shall come back to this point in Section 3.2 and in Vol. 6.)

Every method, regardless of its scope, can be construed as a device transforming problems, together with the data relevant to them, into solutions. See Figure 7.4. In this context we understand by 'data' not only the givens of the problem but also all the components of the body of relevant available knowledge, including ontological and epistemological principles. And by 'solution' we understand the set of solutions of the given problem; thus the two roots of a second degree equation constitute the solution to the latter.

(A method M may be construed as a partial function from the cartesian product $2^B \times \Pi$, where B is a portion of background knowledge and Π a set of problems, into the set S of solutions. The solution to $\pi \in \Pi$ is $s = M(b, \pi)$, where $b \in 2^B$. We take the power set 2^B of B, rather than B, because every method calls for a whole set of bits of knowledge. The *M-soluble problems*, or problems soluble with the help of M, constitute a subset Π_0 of Π, such that the restriction $M_0 = M | \Pi_0$ is a bijection—i.e. one problem—one solution and conversely. In principle there is at least one method for every problem, and different methods may yield different solutions. Two methods are *equivalent* if they yield the same results. But of course they do so by introducing different intermediate steps between problem and solution. In the following diagram two alternative intermediate steps are shown.

In this commutative diagram $M_1 = F_1 \circ G_1$, $M_2 = F_2 \circ G_2$, and $M_1(\pi) =$

$$\begin{tikzcd}
& I_1 \arrow[dr, "F_1"] & \\
2^B \arrow[ur, "G_1"] \arrow[rr, "M_1", "M_2"'] \arrow[dr, "G_2"'] & & S \\
& I_2 \arrow[ur, "F_2"'] &
\end{tikzcd}$$

$F_1(G_1(\pi)) = F_2(G_2(\pi)) = M_2(\pi)$. A *tactics* is thus an ordered pair $\langle F, G \rangle$ such that $F \circ G = M$.)

Knowledge of methods is called *methodological*, and its complement *substantive*, knowledge. These two are distinguishable but not detachable, for the former is only a means for the latter, and it cannot be improved except with the help of the latter. (Imagine trying to design a new particle detector without knowing anything about particles.) Yet from the 17th century on there have been enthusiasts and even cultists of method, claiming that there is a method for every activity, and if only one hits on the right method he will perform to perfection and without any need for substantive knowledge. (An early and delightful exposition of this creed, methodolatry, was Bacon's *New Atlantis*.)

Methodolatry was badly shaken by the realization that (*a*) we know how to do a number of things without being able to formulate any explicit rules for doing them, and (*b*) we know a number of methods that often fail. Loving and building theories exemplify the former, and weighing and making statistical inferences are instances of the latter. (Recall the distinction between explicit and tacit knowledge: Ch. 2, Section 2.2.)

Methodoclasm, the opposite of methodolatry, holds either that there are no methods at all or that they are all equally fallible. This belief is even more wrong than methodolatry, for (*a*) actually there are a number of methods that work reasonably well, and (*b*) all methods, or at least the ways (tactics) of applying them, can be improved. Hence whereas the study of methodology does not produce instant scientists or technologies, its ignorance does not help discover and correct errors.

The advice often given by methodoclasts, namely "Study the history of knowledge instead of methodology", is only a partial answer, particularly because every historian writes from his own perspective, which may be more or less perceptive and tied to a more or less adequate methodology.

We do need methodology in addition to descriptive epistemology and the history of science and technology, if only because (*a*) people do not always think correctly—e.g. most of us tend to disregard unfavorable evidence and write off criticism, (*b*) there are thousands of effective techniques for analyzing statements and arguments, as well as for producing and checking data—not to speak of the uncounted methods for handling problems in pure and applied mathematics, and (*c*) the scientific method embodies certain rules of conduct that constitute the core of the ethics of science, and make it possible to search for truth.

Finally, a word on the value of methods. Each method is valuable to some extent (possibly nil) for solving problems of a given kind and with a certain aim. Change problem or goal and the method may cease to work or it may work better. (In other words, the valuation function for methods is defined not only on the set of methods but also on the sets of problems and goals.) Consequently when comparing methods we must take all three aspects into account. Thus it may happen that one method for solving a given problem is better than a second method with regard to accuracy but worse with regard to speed; or that one method is more general but also less simple than another. For example, the scientific method is the most general and accurate of all methods of inquiry, but it is also the one demanding the largest investment in knowledge and equipment.

3. APPROACH AND RESEARCH

3.1. *Approach*

An approach is—roughly and metaphorically—a way of looking at things or at ideas, and therefore also of handling problems concerning things or ideas. Take for instance the matter of pest control. The chemical approach to it results in spraying pesticides regardless of any undesirable long term consequences: it is determined by a narrow viewpoint and short term goals (saving this year's crop or favoring the chemical company). On the other hand the biological approach to pest control is determined by a broad view as well as long term goals, and it results in trying multiple biological means, such as protecting or spreading certain species-specific predators (e.g. Lady Bugs), and developing pest-resistant strains.

We shall distinguish seven broad approaches to the study and handling of things or ideas: the vulgar, doctrinaire, mathematical, scientific, applied, technological, and humanistic. The *vulgar approach* rests on ordinary knowledge, tackles both cognitive and practical problems, employs daily life procedures (in particular authority and trial and error), and is mainly interested in practical results. The *doctrinaire approach* rests on some ideology, tackles both cognitive and practical problems, resorts to authority, criticism and argument, and is mainly interested in practical results (including personal salvation and the preservation of the ideology).

On the other hand the *mathematical approach* is characterized by a formal basis (logic and mathematics), formal problems, conceptual methods, and the finding of patterns as well as the erection of formal theories. The approach of *basic science* rests on a fund of mathematical and scientific knowledge, as well as on the scientific world view, it deals with cognitive problems, employs scientific methods (in particular *the* scientific method), and aims ultimately at understanding and forecasting facts with the help of laws. The approach of *applied science* shares the basis and methods of basic science, but tackles only special cognitive problems, and aims at supplying the cognitive basis of technology. The *technological approach* is like that of applied science, but its basis includes also the fund of technological knowledge, and its aim is the control of natural systems as well as the design of artificial ones. Finally, the *humanistic* approach is based on the body of knowledge concerning human culture, it handles problems concerning intellectual and artistic culture, uses conceptual methods, and aims at understanding its referents.

In general, an approach may be construed as a body B of *background knowledge* together with a set P of problems (*problematics*), a set A of *aims*, and a set M of methods (*methodics*):

$$\mathscr{A} = \langle B, P, A, M \rangle .$$

Needless to say, every component of this quadruple is to be taken at a given time, and so is a (time dependent) collection rather than a fixed set. (Note that an approach is a projection of a conceptual framework as defined in Ch. 2, Section 4.1.)

If the background knowledge is restricted to the general philosophical framework or world view, we obtain three main types of approach: the atomistic (or individualistic or analytic), the holistic (or synthetic), and the systemic (or analytico-synthetic). The *atomistic approach* rests on an atomistic ontology, according to which the world is an aggregate of units of a few kinds, and a reductionistic epistemology, according to which the knowledge of the composition of a whole is both necessary and sufficient to know the whole. The goals of atomism are the same as those of science, and the atomistic methodics boils down to analysis into components (or the top-down method). The *holistic approach* rests on a holistic or organismic ontology, according to which the world is an organic whole that may be decomposed into large partial wholes that are not further decomposable, and an intuitionist epistemology according to which such wholes must be accepted and grasped as such rather than analyzed and tampered with. The goal of holism is to emphasize and conserve wholeness and emergence, and its method is intuition. (Actually intuition is nonmethodical: recall Section 2.1.) Finally, the *systemic approach* rests on a systemic (or systems-theoretic) ontology according to which the world is a system composed of subsystems, and an epistemology that combines realism with rationalism. Its objectives, like those of science and technology, are to understand, predict, and control. And its methodics includes analysis as well as synthesis, generalizing and systematizing, and empirical testing. (For details see Bunge, 1977a, 1977b, 1977f).

Each approach can handle only certain problems: every approach is partly characterized by its own problematics. Thus the atomistic approach can handle only questions regarding individual behavior: since it does not recognize wholes with emergent properties, it sees no point in looking for patterns of global behavior, i.e. for laws of systems as wholes or units. Likewise the problem system of holism is limited: it is hardly interested in finding out, say, the atomic composition of a molecule, or the mechanisms whereby a group of people form or dissolve a society. By contrast, systemism, or the systems approach, keeps the positive aspects of atomism and holism, and it is the approach compatible with the scientific approach. Actually it differs from the latter only in that it presupposes far less substantive and methodological knowledge.

The advantages of the systemic approach are best appreciated by comparison with its rivals. Whereas systemism treats every system as a whole composed of individuals held together by bonds (part of the system structure) and immersed in an environment, atomism neglects both the environment and the structure; holism neglects the composition of both system and environment, and does not analyze the structure; environmentalism focuses on the environment at the expense of both composition and structure; and structuralism neglects both the composition and the environment, not to speak of the history of the system. Only systemism does justice to all three aspects and their changes.

Failure to adopt the systemic approach to the study or design of systems is bound to result in failure to solve the problems concerned or, worse, in creating unnecessary new problems. For example, a marine biologist studying the migration patterns of sea lions will not make much headway unless he also studies the migration patterns of their prey. An evolutionary biologist studying the foraging abilities of honeybees will do well to study at the same time the evolution of the flowering plants preferred by bees. The archaeologist and the historian of technology cannot restrict their attention to artifacts, but must try and embed them (conjecturally) in the society that used them. The epistemologist is bound to be superficial or even utterly mistaken unless he keeps in mind that, although cognition is a brain process, it continues a tradition and it is now stimulated, now inhibited by social relations. The engineer will not succeed in designing useful large scale projects, such as dams and transportation systems, unless he cooperates with social scientists. The statesman will not introduce useful new legislation unless he realizes that every subsystem of society (the economy, the culture, and the polity) has strong links with the other subsystems of it. Although all this sounds obvious, some of it has been discovered only in recent years, and some is still to be learned by many specialists. Indeed, one may speak of the systemic *revolution*, or of the systemic approach as a new *paragon* (or paradigm). However, for the most part it is a silent revolution with but a few outspoken paladins.

In many cases, particularly in practical matters, the systemic approach violates some of the artificial frontiers between the various fields of inquiry. For example, the so-called new archaeology, that employs the systemic approach, merges history, sociology, and anthropology (cf. Sherratt (Ed.), 1980). And the new technology, that takes into account the social dimension of technical systems, is closely associated with the social sciences (see the journal *Technology in Society*). In all such cases unidisciplinarity is given up in favor of *interdisciplinarity*. In the latter approach the variables typical of a number of disciplines are not only taken into account but also interrelated: the resulting model is neither a single picture nor a mosaic of pictures but, to continue the metaphor, a sculpture. By contrast, the *multidisciplinary* approach consists in the simultaneous application of a number of unidisciplinary approaches: the result is a mosaic of pictures or models. (For an alternative analysis of these and other concepts see Jantsch (1972).) In any event, the problems concerning multilevel systems—and all problems concerning humans are of this kind—are not inherently unidisciplinary, multidisciplinary, or interdisciplinary: only our approaches to such problems can be uni, multi, or interdisciplinary. See Figure 7.5. We shall return to this theme in Ch. 14, Section 3.

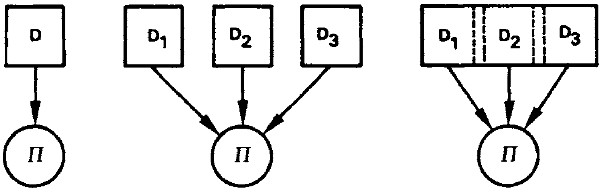


Fig. 7.5. Uni, multi and interdisciplinary approaches to a set Π of problems.

3.2. *Research*

We may speculate that understanding a problem is identical with the activity of certain psychons which, in turn, activate further psychons likely to do the job (Bindra, 1976). Thus if asked to multiply 13 by 67 the subject may choose to do so mentally, with paper and pencil, or with a calculator. In each case different psychons are likely to be activated: which ones depends on the subject's talent and experience, as well as on the instructions he has received. Likewise with problems that require research. Thus if asked to find the law of a certain process, the subject may—depending on his knowledge, intelligence, habits, opportunities, etc.—initiate a bibliographic search, or design an experiment, or sit down and speculate hoping to guess the law on the strength of a few clues such as data and analogies. In each case different psychons are likely to be activated sequentially. As the research proceeds some entirely new concatenations may occur. And, whereas certain psychons will interfere constructively with those in charge of the process, others will interfere destructively, resulting in sidetracking or halting.

Work on a routine problem may be assumed to consist in the sequential activation of a fixed set of psychons. (Analogy: groove or orbit.) Thus when psychon 1 formulates the problem, it activates (with probability near 1) psychon 2, and so on, until the target psychon—the solution—is reached. On the other hand, if the problem is new to the subject, then instead of a linear sequence or chain, a zig-zag process occurs, in which the connections between the various psychons are probabilistic rather than causal, and where the target psychon may never be activated, either because of the subject's personal failure or because the problem has no solution.

Schematically, the routine problem solving process looks like this: *Formulation–search–solution–checking*. The trademarks of this process are definiteness of formulation and near certainty of solution. On the other hand the research problem solving process looks like this: *Raw problem–choice of approach–formulation of problem within this approach–search–research plan–solution candidate–checking–evaluation.* The trademarks of this process are indefiniteness of formulation and uncertainty of outcome. (Nobel laureate Peter Kapitza (1979): "The main attraction of scientific work is that it leads to problems, the solutions of which cannot be foreseen.") In most cases the outcome proves unsatisfactory, and the process has to be repeated either partially or *da capo*, from the very choice of approach or at least from the formulation of the problem within the given approach.

There are no cut-and-dried rules, e.g. algorithms, for performing most of the operations involved in a research problem solving process. At most there are heuristic maxims and methodological rules that help design research plans. Original research, i.e. work on research problems, is zigzagging, not straight; it is chancy, not certain; and it calls for direct cooperation (e.g. of technicians or consultants) or indirect cooperation (e.g. bibliography or referees). Therefore the idea of designing a computer program embodying a "general problem solver" (Newell and Simon, 1972) is wrong-headed.

The most one can do to expedite research is (*a*) to remove obstacles (administrative, ideological, etc.); (*b*) to supply human and material resources, and (*c*) to bring to light some strategies and tactics employed in successful research, and try and clarify, justify, and systematize them—i.e. to cultivate methodology. The strategies of inquiry are the general research methods, such as the scientific method and the methods of random sampling and of mathematical modeling. As for the inquiry tactics, they are of two kinds: techniques (special methods) and heuristics. Whereas techniques are reliable procedures, heuristics are suggestions that may or may not work—such as looking for analogies with problems of known solutions, or trying to generalize special results.

The most general strategy for handling research problems, be it in science, technology, or the humanities, was indicated a while ago, but it bears repetition. It consists in the following sequence of operations:

1. *Spotting the problem(s).*
2. *Choice of approach* (which includes background knowledge, methodics, and goal).
3. *Formulation of problem* within chosen approach (i.e. transformation of raw initial problem into well defined problem).

4. *Search for existing relevant information* (i.e. scanning of background knowledge stored in theories, tables, data banks, etc.).

5. *Design of a research plan.*

6. *Implementation of plan* (conceiving hypotheses, building theories, making computations, performing measurements, designing techniques or artifacts, etc.).

7. *Trying* (checking) the outcome of previous step.

8. *Evaluation* of candidate solution (idea, data, artifact, or what have you) in the light of the background knowledge as well as the outcome of the previous step.

9. If solution is found acceptable, *extend or revise* background knowledge.

10. Otherwise *start research process again*, either replicating it or modifying some components (e.g. hypotheses or methods).

Note again the difference between a routine and an original research project. The former ends up by answering questions, the latter poses further questions even when it attains answers. (In other words, whereas routine problems are information sinks, original problems are seminal. Yet not all sources are sizable, and not all sinks are dead ends: some findings of routine research are used in further research even when they pose no new problems. Most data are of this kind.) Routine research is closed, original research open ended. The former is straight, the latter tortuous: it rarely attains its goal in a direct manner, usually it zig-zags, sometimes wanders off, and at other times derails. What one reads in the published reports is an expurgated and embellished version—which is just as well, for a faithful chronicle would be too long, convoluted, incoherent, full of uninteresting detail; and most mistakes are unenlightening except to those who make them.

(Routine research can be construed as a partial function from problems to sets of findings, or $R:\Pi \rightarrow 2^F$, where F is the set of all old and new, actual and potential, outcomes of research in the same field as Π. F does not include the null finding 0. We say that research into problem $\pi \in \Pi$ ensues in a knowledge gain $R(\pi) = f \in F$ provided $f \neq 0$. In the case of routine research the background knowledge, the methods and the aim of research are fixed and taken for granted, so there is no need to include them in the domain of R. On the other hand original research can vary these items, and results not only in findings but also in further problems. Therefore it may be construed as a partial function from quadruples ⟨bits of knowledge, problem, method, aim⟩ to couples ⟨set of findings, problem⟩, or $R:2^B \times \Pi \times M \times A \rightarrow 2^F \times \Pi$. See Figure 7.6.)

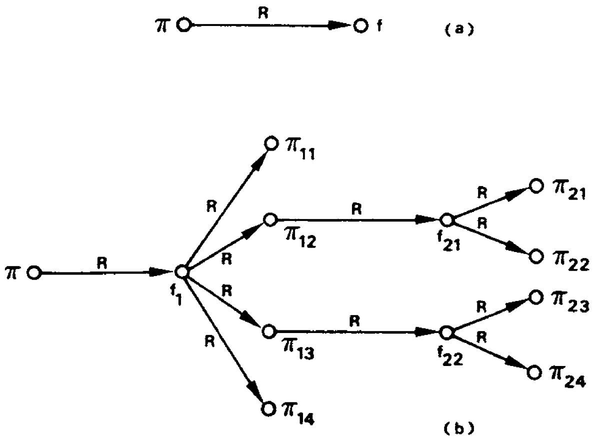


Fig. 7.6. (*a*) Routine research transforms problems into findings. (*b*) Original research transforms problems into findings or into further problems. Some problems are not investigated at all.

(The two functions, routine research and original research, may be thought to be embodied in sequential machines. Routine problem solvers may be construed as sequential machines with fixed next state and output functions: every input takes them to another state in a fixed state space, and induces an output from a fixed set of possible outputs. On the other hand original problem solvers may be regarded as probabilistic and self-organizing sequential machines, i.e. machines with variable next state and output functions, and moreover probabilistic functions. However, such devices are not machines proper, for there is no way to design or program a machine to enlarge its own state space, or create its own next state and output functions, but to specification—in which case it is a routine not an original problem solver.)

Our description of the general strategy for handling research problems tells us *what* is to be done but not *how* it is to be done. The how is a matter for *heuristics*, or the art of problem handling. The following heuristic suggestions may help go through the research problem solving process. (See Pólya, 1954, 1957; Bunge 1967a.)

1. *Search for a good problem*

Scan the background knowledge for a gap that you feel is worth filling and you might fill yourself.

Consult with coworkers or assistants, colleagues or teachers.

Pay no attention to pessimistic advice if you feel sure you have got hold of a good problem and can afford to fail in the attempt to solve it.

2. *Choose the right approach*

Start by locating the problem in a discipline or a group of disciplines.

Adopt an approach—basic or applied, atomistic or systemic, etc.

Determine the minimal background knowledge required to handle the problem.

Find out the recent history of the problem or that of its predecessors.

3. *Locate the problem*

Determine whether the problem is one of original research or routine, substantive or methodological, conceptual or empirical.

4. *State the problem clearly*

Minimize the vagueness of concepts and the ambiguity of signs.

Select adequate symbols—as simple, suggestive, and standard as possible.

Draw sketches or diagrams, build preliminary scale models, or make preliminary calculations.

5. *Make a further study of the problem*

Identify the premises and the unknowns.

Unearth the most important presuppositions and examine any that look suspicious. (As a result you may decide to switch to another problem, e.g. the problem of reconstructing such presuppositions.)

Eliminate the redundant information ("noise").

Compress and simplify the data.

Formulate and organize your preliminary hypotheses and rules of procedure.

Introduce simplifying assumptions or devices.

Look for similar solved problems or subproblems (components of the given problem).

Vary the constituents or the formulation of the problem, in an attempt to transform the given problem into a homologous problem in an adjoining field.

6. *Design a research project (plan, program)*

Reexamine and reorganize your preliminary hypotheses and rules of procedure.

Organize the steps of the proposed research process according to their logical priority (A logically presupposes B), epistemological priority (Knowledge of A presupposes that of B), or pragmatic priority (Doing A requires doing B first).

Choose the technique adequate to the nature of the problem and the kind of solution wanted.

Estimate the possible advantages and shortcomings of the various available techniques, if there are any.

In case no technique is available, state the strategic problem of devising one, and attack this one first.

Choose the materials—drugs, animals, locations, etc.—and the instruments—centrifuges, thermometers, computers, etc.—likely to be used.

Choose the coworkers, assistants, and technicians.

Secure the financial means.

Make provisions for the unexpected—in particular for changes in plan or even in problem.

Be prepared to have your project rejected for being either too poor or too good.

The preceding operations constitute the preliminary stage of the problem-solving process. Although they do not solve the problem, they do prepare it for attack, and they gather all the means likely to be required to solve it—or to show that the problem has no solution with such means. Actually the preliminary stage is usually the most difficult of all: a correct statement and a good strategy is half the solution. No wonder then that such preparation may consume as much time and ingenuity as the implementation of the research project.

The question of formulation is so important that we must insist on it, the more so since most problems in everyday life can be solved without formulating them explicitly, for we have learned well-tried routines to solve them. Other problems must be formulated explicitly if only because they admit of alternative formulations and therefore alternative solutions. (Better: a raw problem is actually a whole set of well formulated problems.) For example, if we wish to find out where two curves meet, we must either draw the curves (obtaining an approximate solution) or solve a certain equation. If we wish to find out whether a given material conducts electricity, we may insert it in an electric circuit and touch it (at our peril) or watch whether the pointer of the ammeter moves—or make assumptions about the composition and structure of the material, and compute its conductivity with the help of quantum mechanics. In science, technology and the humanities all problems have to be formulated explicitly and with maximum clarity. Such formulation (or refinement of the raw initial problem) is seldom unique: it depends on the state of the art as well as on the outlook, experience, talent, and aims of the investigator. (For the importance of such personal factors see Polanyi (1958) and Holton (1973).)

The preliminary study of a problem culminates in a *research project*, plan or program. Any such project looks, in a nutshell, like this: Given a certain background knowledge b, and an approach a, attack problem π on the basis of programmatic hypothesis h, aiming at goal g, and using means (among them methods) m (available or to be produced). In short, a research project is a sextuple $p = \langle b, a, \pi, h, g, m \rangle$. *Example 1*: Given the available body of sociological knowledge, and ordinary mathematics, and adopting the systemic approach, build a comprehensive theory of social relations, on the basis of the programmatic hypothesis that all social relations are either interpersonal or intersystemic; the goal is to explain social behavior, and the means the standard method (or style) of model construction in the "hard" sciences. *Example 2*: Given the available body of neurophysiological and psychological knowledge, and adopting the bottom-up approach, identify the neuronal systems involved in thinking abstract ideas, on the basis of the programmatic hypothesis that thinking is the specific function of certain plastic neuronal systems; the goal is to furnish a deep explanation of thinking, and the means all the paraphernalia of neuroscience, experimental psychology, and mathematics.

Every independent investigator has some research project or other, whether original with him or supplied by somebody else—laboratory director, thesis advisor, senior colleague, etc. If vast and ambitious, a research plan is likely to be shared by, or passed on to, other investigators. And, on important issues, several competing research projects may be at work contemporaneously. Hence every field of active inquiry may be viewed as a changing bundle of more or less interdependent research projects in progress. (There is no harm in duplication: on the contrary, duplication is necessary for control.)

The way to become an investigator, i.e. to master the craft of research, is not by becoming acquainted with instances of good science—e.g. by reading history of science—or with methodology—e.g. by reading this book. Both help but neither is a substitute for participating in a research project at the knee of some master—much the same way that craftsmen and artists learned their trade in the Middle Ages and the Renaissance. The reason is that learning to do research is not just acquiring some information but rather acquiring certain habits, attitudes, and aspirations—all of which are a matter of know-how.

General methodology cannot go much further into the matter of handling problems. Any more specific suggestions would take us into specific branches of research. For example, a useful heuristic rule in applied mathematics is: "If a term of the form $A_x B_y$ occurs in a calculation, try adding $-A_y B_x$ to it, for the result is the z-component of the vector product of A and B". Biologists use Cuvier's rule: "If two organisms (or fossils thereof) exhibit two homologous traits, look for more". Historians make use (and often abuse) of the rule: "Look for precursors of every novelty". And doctrinaires abide by the rule: "Distrust all new ideas or behavior patterns". There are hundreds of special heuristic rules of this kind, which methodology has yet to gather and study.

4. ANALYSIS OF PROBLEMS

4.1. *Logic and Semantics of Problems*

We proceed now to a logical and semantical analysis of problems. Ordinary logic and semantics do not account for them because they are not propositions and therefore can be neither true nor false. Nor can linguistics help, for problems can be expressed either by interrogatives or by commands or even by sentences. Thus 'What was that?', 'Find out what that was!', and 'You are being told to find out what that was' express the same problem. The only reasonable approach to the analysis of problems is epistemological, i.e. to regard them as the starters of inquiry processes. This is the approach adopted in an earlier work (Bunge, 1967a, Ch. 4), which we shall now review and complete.

To begin with, problems do not come in isolation but as components of approaches. Recall that we defined an *approach* as a quadruple ⟨*Background knowledge, Problematics, Aims, Methodics*⟩ (Section 3.1). But of course not every problem stands out against the entire background knowledge: only a part of the latter is required to pose and solve a problem. We call this part the *setting* or *context* of the problem. More precisely, the setting or context of a problem in a given approach is the collection of components of this approach that are related in some way (logically, semantically or pragmatically) to the concepts occurring in the formulation of the problem. For example, one suitable setting for investigating the mechanism of problem finding and solving is neuropsychology; another is the sociology of knowledge; a third is the history of knowledge.

In principle every problem can be approached in more than one manner, and so it has more than one setting. Thus the problem of how to eliminate poverty belongs to, or arises in, ethics, ideology, politics, and applied social science. Because a problem may belong to different contexts, it may receive different solutions. For example, drug abuse is seen in a biological setting as an instance of addiction, in a medical setting as a matter of cure, in a legal setting as one of legislation and law enforcement, and so on. Each setting suggests the way to handle the problem.

Although a problem can be approached in alternative ways, these are not always equally suitable. In some cases one of the approaches is preferable to the others, and in other cases the merging of several approaches is desirable. For example, the problem of the origin of life is approached differently in science and in religion. The former approach invites making conjectures and experiments on the self-assembly of biomolecules, organelles, and eventually cells; on the other hand in many religious approaches the matter is one of faith in some dogma or legend about a divine *fiat*. In this case, then, the scientific approach will be preferred only by those who prefer truth by research to truth by revelation. In other cases different approaches are equally legitimate but partial, and must therefore be joined. Any social problem, such as poverty, ignorance, or delinquency, is a case in point: none of them can be solved within a single setting, but every one of them calls for interdisciplinary studies and measures.

In some cases the context of a problem must be shifted. Thus when a mathematician is summoned to help with a scientific or technological problem, he treats it as a problem in pure mathematics, i.e. he drops the factual interpretation of the concepts involved. For example, time will be for him just a real variable, and an equation of motion just one more equation to be investigated on its own mathematical merits. Likewise the biochemist moves biological problems over to chemistry, by reducing them to chemical reactions and studying them *in vitro*. A similar shift of context occurs when science is applied. For example, the phytotechnician must "translate" the results of the genetics laboratory into cultivation methods, crop yields, and so on. All these are cases of problem relocation or *recontextualization*.

In sum, every problem arises in some context or other—or, equivalently, it is a component of some approach or other. And it can be moved to some other context, not necessarily to advantage. There are no problems out of context, but the various possible approaches to a problem may not be equally suitable for a given purpose. And sometimes a systemic approach, involving the joining of a number of contexts, is required.

Our next task is to analyze the internal structure of problems. We postulate that every problem has three constituents: presupposition(s), generator, and solution (Bunge, 1967a, Ch. 4). Consider the problem "What thinks?", i.e. "What does the thinking?". It presupposes (takes for granted, does not question, assumes the premise of) the existence of some entity, be it the brain or the mind, that does the thinking. It is generated by the predicate "thinks", for the unknown to be uncovered occurs in the underlying formula "x thinks", or "Tx" for short. And the problem induces research resulting in a solution of the form "b thinks", or "Tb", where 'b' names a definite individual—someone's brain or mind. In other words our problem is "Which is the x such that x thinks?", which we symbolize '$(?x)Tx$'. The presupposition of this problem is "Something thinks", or "$(\exists x)Tx$", the generator is "T", and the solution "Tb", or simply "b". Logically, then, we have the sequence: *presupposition–generator–solution.* We may identify each class of problems with such a sequence.

In general, every problem is posed against a certain background of antecedent knowledge. This background contains the context or setting of the problem, and in particular its presuppositions. The latter are the propositions (data or hypotheses) that are somehow involved though not questioned in the inquiry prompted by the problem, although they may become the subject of alternative inquiries. The problem may be regarded as generated by a set of open formulas (what used to be called 'propositional functions') exhibiting gaps in knowledge, i.e. unknowns, that are to be filled by research. The formulation of the problem results, formally, from applying the operator "?" one or more times to such formulas. However, "?" is an epistemological operator, not a logical one: it means "search for" without even intimating how to conduct the search. Finally, research on a soluble problem yields eventually another set of constructs—the solution to the problem—which, when substituted into the generator, turn the latter into a set of statements that can be assigned a definite truth value.

The notion of presupposition of a problem occurs in the following definition of the important concept of meaningfulness of problems. Let C be a context included in a given body B of background or antecedent knowledge. We stipulate that problem π is *well conceived* in C only when the following conditions are jointly fulfilled: (i) π belongs to some approach $\langle B, P, A, M\rangle$ including (actual or possible) methods in M capable of handling π; (ii) all the concepts occurring in π occur also in C, hence in B, and (iii) the presuppositions of π are compatible with C. (We require the latter to rule out crackpot questions, and do not require compatibility with the entire background B, to allow for revolutionary questions.) Otherwise π is *ill conceived* (or *makes no sense*) in C. Thus the problem "When did God create man?" is ill conceived in modern biology, which does not contain any concept of divinity, and asserts that man was not created but evolved from ape-like ancestors. In general, theological questions are meaningless in the context of science and conversely. (More in Ch. 14, Section 4.2.)

At first sight such statements as "Where is p?", "When does p happen?", "Is p true?" and "Is p valuable?" do not fit the above schema, for they seem to be generated by p itself, which is now a thing, now an event, now a proposition, now any object. Yet such problems, when restated explicitly. do prove to fit our schema. Thus "Where is p?" is best reformulated as "What are the values of the coordinates of p relative to such and such reference frame?", or $(?v)(V(p) = v)$ for short, where in this case v is a triple of real numbers. Likewise "Is p true?" is best construed as "What is the truth value of p relative to such and such body of knowledge?", or $(?v)(V(p) = v)$, where now v is any real number included in the interval $[0, 1]$. Every one of the above problems presupposes something or other—e.g. that every thing is at some place (relative to some frame), or that every proposition can be assigned a truth value (relative to some body of knowledge). And they are all questions about the value of some individual variable or other, even if the variable is hidden at first sight.

Not all problems concern individual variables or unknowns: in some cases the blanks to be filled are predicate variables. For example, the problem "What is b?" directs us to search for the cluster of properties P, so far unknown, possessed by b. This problem may be symbolized by '$(?P)Pb$', which points to the unknown P, a conjunction of attributes. Questions asking for the value(s) of one or more individual variables shall be called *individual problems*, whereas questions asking for the value(s) of one or more predicate variables shall be called *functional problems*. We assume that every elementary problem is of either of these kinds, and every complex problem a combination of problems of either kind. We shall deal below with simple and complex problems.

What about problems of the form "Does b have the property B?", in which no variable is in sight? The variable is hidden: it is the truth value of the proposition "b has the property B". In other words, the given problem is actually of the form $(?v)(V(Bb) = v)$, where V is a truth valuation function. The problem consists in finding out the precise value of v.

Similarly, all questions concerning existence can be regarded as problems about the truth values of the corresponding existence statements. Thus "Are there free quarks?" can be restated as "Is it true that there are free quarks?", which in turn is rendered explicit by "What is the truth value of the proposition that there are free quarks?". Similarly with questions concerning universality. For example, "Does every thing change?" is equivalent to "Is it true that every thing changes?", which in turn may be restated as "What is the truth value of the thesis that every thing changes?". The moral we extract from examining the above cases, characterized by the seeming absence of variables, is this: Do not allow language to mislead you—*cherchez la variable.* Note also that the question mark bears always on some unknown or variable of either the individual or the predicate type. (Furthermore "?" does not bind the variable on which it acts. Only the solution should be free from unbound variables, i.e. it should be of the general from Bb, where B is an n-ary predicate and b an n-tuple of individuals.)

(The question mark, which so far we have left undefined, can be partially defined as follows. We start by abstracting from the variable type, writing an arbitrary elementary problem in the form: $\pi(v) = (?v)G(v)$, where v is the only variable and G the generator. Our new formula subsumes both elementary problems of the individual type, i.e. $(?x)Bx$ with a fixed B, and elementary problems of the functional type, i.e. $(?X)Xb$, where b is a fixed individual. Although these problems are very different, they are equivalent in the sense that they have the same solution, namely Bb. Further, let $R{:}\,\Pi \rightarrow 2^F$ be the routine research function introduced in Section 3.2, that represents an inquiry process as a mapping from problems in Π to sets of findings in F. We stipulate the partial implicit definition

$$R(\pi) = R[(?v)G(v)] = Bb,$$

i.e. research on a problem yields the solution to the latter.)

(The generalization to a research or original problem is straightforward. In this case, as we saw in Section 3.2, the outcome of research is in general an ordered couple $\langle$Finding, New problem$\rangle$—where the finding is occasionally that the problem is not soluble with the indicated means. The case where no new problem arises is $\langle$Finding, Null problem$\rangle$, where *Null problem* $= (?x)Ox$ or $(?X)Xo$, and in turn O is the null predicate—attributable to no individual—and o the null individual—to which no predicate except O is attributable. Likewise, if the only result is a new problem, we write: $\langle$Null finding, New problem$\rangle$, where the null finding is the one not in 2^F.)

The individual/functional dichotomy of problems cuts across many other classifications. For example, Aristotle distinguished between *what*-problems, or questions of fact, and *whether*-problems, or dialectical (logical) problems; and Pólya (1957) between *problems to find* and *problems to prove*. We shall not enter into this matter but shall comment instead on four kinds of problem that are of particular philosophical interest: inverse, why, what-if, and existence problems.

The *inverse* or *dual* of a problem is one where the data of the latter are the unknowns of the former and conversely. Thus if computing a value of a function is a direct problem, then calculating the argument(s) from the value is the corresponding inverse problem. Here is a small random sample of pairs direct-inverse problems. Given an equation, find its solutions; given a function, find the equation it solves. Given a theory, check a theorem in it; given the latter, find whether it is a theorem of a given theory (i.e. solve the completeness problem for the latter). Given the mass of a particle and the force(s) acting on it, find the particle trajectory; given the mass and the trajectory, find the force(s).

The last mentioned example illustrates a very common situation in all fields of inquiry, that of analysis and synthesis. The *analysis* problem is: Given a system (i.e. knowing or assuming its composition, environment, and structure), find its behavior. The inverse problem is that of *synthesis*: Given a behavior, find or design a system that realizes it. For example, find a kind of molecule or material, machine or cultivar, social system or teaching method, that will behave in a prescribed way or achieve given goals. Typically, the problems of technology and of social action are synthesis problems; with luck they are the inverses of scientific problems with known solutions. As a rule inverse problems are far harder than the corresponding direct problems, if only because they usually have multiple solutions.

Thus think of the following problem pairs. (*a*) Direct problem: Given a body, a screen, and a projection method, find the projection or shadow of the body on the screen; inverse problem: given the projection on the screen (e.g. a photograph) and the projection method, reconstruct the body. (*b*) Direct problem: Given a social policy, find the state of the society resulting from its application; inverse problem: given a desirable state of society, find a social policy likely to achieve it. (*c*) Direct problem: $(?x)(\sin b = x)$; inverse problem: $(?x)(\sin x = b)$ or, in an equivalent notation, $(?x)(sin^{-1} b = x)$. In this case the direct problem has a single-membered solution, for *sin* is a function; but, because *sin* is not a 1:1 function, the inverse problem has a multiple—in fact an infinitely-membered solution. (If a is a solution to it, so is $a + 2n\pi$, where n is an arbitrary integer.) So much for the fascinating and ill-studied problem of inverse problems.

Why-problems are of the form "$(?p)\ (p \vdash q)$", i.e. "What premises p entail the given q?". This is the kind of problem that the theory builder faces: his givens are low level generalizations, and his desideratum is to invent high level generalizations entailing the low level ones. (More in Ch. 10.) *Why*-problems are the inverses of *what-if* problems, which have the form $(?q)\ (p \vdash q)$, or "What are the consequences of p?", or "What would happen if p were the case?". These are dialectical problems, or problems to prove. What-if questions fall into two kinds: those with false and those with true presuppositions. "What if we could live without eating?" illustrates the former kind. Such questions generate fiction rather than knowledge. On the other hand a problem such as "What if the nebular red shift were due to some cause other than the expansion of the universe?" violates no known law of physics and is fruitful, for it suggests inquiring into the possibility of a cosmology without the Big-Bang.

Existence questions are among the most difficult of all. They can be formulated thus: "Are there objects of kind K?", or, more generally, "Are objects of kind K possible?" where K is determined by fairly exact predicates. That question can in turn be restated as "What is the truth value of the statement that K is or could be nonempty?". They are of two main types: formal (e.g. existence of solutions to a given equation) and factual (e.g. existence of a planet). Consequently they are to be investigated conceptually in the former case, and empirically as well as conceptually in the second. (Recall that the expression 'exists' is ambiguous.) In science and technology existence (or possibility) problems are not whimsical: there is always some ground to conjecture that something exists or may be brought into existence. And answering such question is prior to asking the more specific question 'How do K's exist?', or 'How are K's possible?'.

Factual existence problems can in turn be classed in various ways, e.g. thus: (*a*) the most general ontological existence problem: 'What kinds of thing are there (really)?'; (*b*) scientific existence problems, of the form "Does the such and such exist?"; (*c*) the general epistemological existence problem, "To what may we legitimately attribute existence?"; (*d*) the most general technological existence problem, "What kinds of thing can we design and bring into existence?"; (*e*) special technological problems, of the form "Can we design and construct such and such things?"; (*f*) the most general ethical existence problem, "What ought we to bring into existence?"; and (*g*) special ethical existence problems, of the form "Ought we to bring such and such into existence?".

The ontological question is wider than the partial scientific existence problems. (It would seem that there is a prior ontological problem to be solved, namely "Why is there something rather than nothing?" However, this problem makes sense only in a non-naturalistic context, for it presupposes that there is a Creator and that it might have been possible for Him not to go to the trouble of making the world. The problem makes no sense in a naturalistic ontology, let alone in science.) The epistemological question has a deceptively simple answer: We can attribute existence to whatever exists really. This answer is unhelpful, so we must settle for a weaker one, namely: We may attribute real existence to whatever we have good empirical, and if possible also theoretical, grounds to conjecture that it exists. (For a definition of real existence see Bunge (1981a).) The general technological question is of course broader than any specific technological problem such as "Are there antidotes to such and such poison?" or "How is that subject best taught to undergraduates?". The same holds for the relation between the general and the special ethical questions. But what is such a relation between a general and a special problem, or between a problem and its subproblems if any? Let us find out.

Consider any two problems. We can form a third problem out of them if we set up the task of solving one *or* the other, and a fourth problem if we propose to solve the one *and* the other. We say that in the first case our compound problem is *disjunctive*, in the second *conjunctive*. The solution to a disjunctive problem is the disjunction of the solutions to its components, whereas the solution to a conjunctive problem is the conjunction of the solutions to its components. And we form the *negate* of a problem if we negate its generator. The solution to the negate of a problem is the negation of the solution to the problem.

The three operations can be combined in a number of ways to form problems of any complexity. Conversely, a complex problem can be analyzed in terms of the disjunction, conjunction, or negation of elementary problems. For example, the algebraic problem $(?x)$ (x is a complex number & $x^2 + x - 2 = 0$) can be rewritten: $(?x)\,[\mathbb{C}x \,\&\, (x-1)(x+2) = 0]$, and thus be decomposed into two elementary problems, namely $(?x)$ $(\mathbb{C}x \,\&\, x - 1 = 0)$ and $(?x)$ $(\mathbb{C}x \,\&\, x + 2 = 0)$. The original problem is thus seen to be of the disjunctive type, and its solution: $(a=1) \vee (b=-2)$. Whereas sometimes it pays to analyze a complex problem into its constituents, at other times it is advantageous to switch from a problem to its negate. Thus to find out whether a given person is celibate we may assume that she is married, and search for her spouse; likewise to find out whether a thing is far away we had better start by exploring the neighborhood.

(The three operations on problems which we have just introduced may be symbolized as follows. Let $\pi(v)=(?v)Gv$ be a problem with generator G and solution $G(a)$. Then the negate of the problem is defined to be

$$-\pi =_{df} \neg G.$$

Call now $\pi(v_1)=(?v_1)G_1(v_1)$ and $\pi(v_2)=(?v_2)G(v_2)$ two problems with solutions $G_1(a)$ and $G_2(b)$ respectively. Then the disjunction and the conjunction of these problems are defined as

$$\pi(v_1) \curlyvee \pi(v_2) =_{df} [G(a,b)=G_1(a) \vee G_2(b)],$$

$$\pi(v_1) \curlywedge \pi(v_2) =_{df} [G(a,b)=G_1(a) \& G_2(b)],$$

respectively, where $G(a,b)$ is the solution to the corresponding binary problem.)

On the basis of the three operations on problems, which we have introduced, more complex problem forms can be produced and analyzed. Also, the very concept of a complex problem can be defined, namely thus: A *complex* (or *compound* or *molecular*) problem is one where either the negation or the conjunction or the disjunction of problems occurs. The definition of the dual concept is obvious: a problem is *elementary* (or *simple* or *atomic*) if it is not complex. Sometimes it pays to increase the complexity of a problem, as when an insoluble problem about real functions is translated into a soluble one about complex functions, and when an economic problem is regarded as just one component of a complex social problem. However, most of the time we wish to simplify problems by analyzing them into disjuncts, conjuncts, or negates. Thus a triatomic problem may be analyzed into a disjunction of conjunctions. (In effecting such analyses we assume the associativity and commutativity laws; we also assume that disjunction distributes over conjunction and conversely.)

Besides the three operations on problems, we introduce three useful relations between problems: those of implication, equivalence, and entailment. We stipulate that a problems *implies* a second problem if the generator of the first implies that of the second, and that two problems are *equivalent* if their generators are. We also stipulate that a problem *entails* another problem if the generator of the former entails that of the second, i.e. if the solution to the second follows from that to the first. Finally, we say that a problem is a *subproblem* (or *special case*) of another problem if the second implies or, in particular, entails the first. For example, every homogeneous equation is a subproblem of any number of inhomogeneous equations, and the problem of minimizing alienation is a subproblem of that of maximizing participation.

We can now lay down rules for the proper formulation of problems:

Rule 1: A problem π is *well formulated* (or *proper*, or *sound*) in a context C iff

(i) π is well conceived (or makes sense) in C (see the definition of this concept at the beginning of this section);

(ii) the generator of π contains as many variables and question marks as unknowns—i.e. the statement of the problem indicates explicitly what is to be sought;

(iii) π is elementary (or atomic) of the individual or the functional type, or is a combination (by means of the operations of disjunction, conjunction, or negation) of elementary problems.

A well formulated problem is determinate or well-defined: if soluble at all with the given means it will have a unique (though possibly many-membered) solution. And, by displaying all the relevant items (the data and unknowns), as well as their relations, it should suggest the type of inquiry needed to solve it. For example, a system of m linear equations in n variables (unknowns) is *determined* if and only if $m = n$, *underdetermined* if $m < n$, and *overdetermined* if $m > n$. The problem is well formulated only in the first case. If it is underdetermined, it has no unambiguous solutions—or, if preferred, it has too many solutions. If it is overdetermined it has no solutions at all (unless some of the equations are redundant). Therefore if the problem is underdetermined we must try and get further equations or eliminate some variables. And if the problem is overdetermined we must try and reduce it to a determinate problem by eliminating $m - n$ equations.

Rule 1 should be helpful to weed out ill-formulated problems but it cannot be expected to lead to formulating good problems in the sense of Section 1.2. It does not even guarantee the formulation of well conceived problems, because it is always difficult to unearth and examine all the relevant presuppositions. Even in a highly formalized theory only those presuppositions that have been identified by the theorist are listed. Except in trivial mathematical cases, every list of presuppositions is likely to be incomplete. Although this incompleteness causes sometimes the problem to be ill-conceived, the very discovery of such incompleteness may be the source of interesting new problems.

So far we have dealt with isolated problems. But in ordinary life and in research fields problems come in clusters, not in isolation. This is particularly true of original research in science and technology, where every problem has some precursor and triggers research ending up in further problems. (This is what is meant by the popular saying 'Everything leads to something else'.) Thus "What is it like?" precedes both "How did it evolve?" and "What is it for?". The precedence in question is epistemological in the case of cognitive problems: we must solve certain problems before being able to attack others. And it is pragmatical in the case of practical problems: we must make certain means available before we may endeavor to attain certain ends.

So, every problem, whether cognitive or practical, belongs to some *branching sequence* of problems, i.e. a partially ordered set. Discovering, planning or manipulating such partial orderings of problems is part of the strategy of research or of action. Such strategy must be sketched, if only to be modified whenever necessary, unless the search or the action are to be haphazard and therefore inefficient. Evidently, the ordering relation is this: a problem *precedes* another if solving the latter presupposes solving the former—i.e. if the solution to the first is used to solve the second.

(In the case of cognitive problems this relation can be elucidated thus. Let C_1 and C_2 be the contexts or settings of problems π_1 and π_2 respectively, and call $G_1(a)$ and $G_2(b)$ their respective solutions. Then $\pi_1 < \pi_2$ iff $G_1(a)$ is in C_2. In particular, if the problem is one of proving, C_2—which includes $G_1(a)$—entails $G_2(b)$. In the case of practical problems we can say that $\pi_1 < \pi_2$ iff the solution to the first problem is among the factual conditions—causes, actions, materials, or what have you—needed to bring about the solution to the second problem. Since practical problems are concerned with the modification of some concrete system or other, problem precedence amounts to this: $\pi_1 < \pi_2$ iff the system has to attain the state associated with the solution of π_1 before it can reach the state associated with the solution to π_2.)

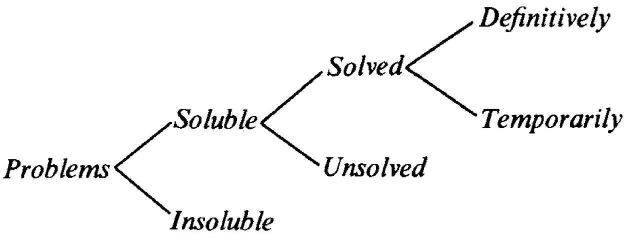
(Actually problem clusters are richer than ordered sets: they are *networks*. In fact, given any problem we can form its negate, and given any two problems we can form their disjunction and their conjunction; besides, problems come in direct-inverse pairs. This suggests defining a *problem system* as a nonempty set of problems, ordered by the relation of problem precedence, together with the operation of problem disjunction, conjunction, negation, and the still unclear operation of problem inversion. The algebraic structure of such a system is yet to be investigated.)

We postulate that *every research problem belongs to some problem system.* A corollary of this assumption is that no research problem can be solved unless some other components of its system are attacked as well—some before, others at the same time. Thus the problem of synthesizing a living cell requires the solution of hundreds of problems in molecular biology, biochemistry, cell biology, evolutionary biology, instrumentation, etc. And the energy crisis cannot be solved by adopting a single measure, for it belongs to a vast problem system composed not only of technological problems but also of economic, political and cultural ones. The mere realization of the systemic character of research problems would facilitate research and expedite social readjustments.

4.2. *Solved, Unsolved, Insoluble*

Not all problems, even if well formulated, are soluble; and those which are can be solved with certain data (or assumptions) and methods but not with others. Solubility is relative to data (or assumptions) and methods rather than absolute. Thus the problem of the quadrature of the circle cannot be solved with ruler and compass but is easily solved analytically; and the problem of communicating with the deaf-mute cannot be solved using oral language but is solved by means of sign language.

In regard to their solubility problems can be classed as follows:



Except for routine problems it is impossible to prove conclusively that a given problem is soluble—short of solving it. On the other hand it is sometimes possible to prove that a problem is insoluble in a certain manner—e.g. exactly, or in a single manner—with the help of certain data (or assumptions) and methods. For example, we cannot measure simultaneously the exact position and momentum of an electron because it does not have them; we cannot build a perpetual motion machine because energy is conserved in a closed system; and we cannot communicate with the dead because the dead happen to be dead.

All of the above are examples of total insolubility or, if preferred, of ill-conceived problems or even pseudoproblems. They are not insoluble for want of knowledge but because the world is the way it is. (For example, we cannot communicate with the dead because only the living brain is capable of thinking, and thought can be communicated only through the senses.) They are not permanent mysteries such as those of religion, which are supposed to be beyond the reach of reason and experience. Nor are they relative mysteries such as most murders, which could be solved with some more inquiry. They are just nonproblems.

The genuine but insoluble problems are those the metamathematicians call 'unsolvable': they can be *proved* to be such. For example, Gödel proved the existence of statements that are not provable within certain theories, namely all those containing arithmetic. He did not exhibit any specimens of such statements but just proved that they exist (formally), so his proof did not affect the working mathematician. Still, it undermined the faith that every mathematical problem can be solved within some theory. In factual science and in technology we cannot prove any problem to be unsolvable. Here we know that, whereas some problems have been solved, if not completely and exactly at least partially and approximately, others remain unsolved or even unstated. In factual science and technology we have no mysteries but only unsolved problems.

Here is a random sample of scientific problems still unsolved, or solved only partially, by 1983. Is the universe spatially finite or infinite? Are there black holes? What are quasars? What is the precise origin of cosmic rays? Are there reliable earthquake predictors? What caused the various glaciations? What is the function of the neutral mutations: important or insignificant? What is the function of "selfish genes" (that do not transmit hereditary traits)? Why are most molecules in organisms L-active? Does bioevolution proceed always gradually or is it punctuated by macromutations? At what point does a variety become a species? Is geographic isolation necessary for speciation? What caused the extinction of three-quarters of the biospecies at the beginning of the Tertiary era about 65 million years ago? Why do we have only half as many genes as frogs do? What is the precise lineage of our species and, in particular, whence, when and how did the most primitive *Homo* emerge? Why do all higher vertebrates sleep? Why do most people retain their most important beliefs in the face of refutation? How do we think? How do we learn to talk? Why are some people more interested in solving problems than in nursing beliefs? What are the factors of social cohesion? Are there natural laws of the economy? Are there laws of history? How did capitalism originate? Is socialism compatible with democracy?

All of the above problems are being investigated at the time of writing, and some of them have given rise to heated controversies. Few scientists believe that there are cognitive problems both well formulated and insoluble, i.e. absolute mysteries: they leave irrationalism to theologians and philosophers. Not that it can be proved that all of the above problems are soluble: this is a matter of faith, not proof. (Solutions, not problems, are subjected to proof or disproof, conclusive or otherwise.) Yet this is no naive (uncritical and groundless) faith but one that can be justified in various manners.

First, there is the past success of inquiry every time it has been conducted scientifically and been allowed the freedom to do so. Second, there is the enormous capacity of the human brain to form and dissolve neuronal connections, to activate itself, to respond to novelty, and to seek it. Third, society needs more cognitive problem finding and solving if it is to face its most pressing problems. (Not that science can tackle practical problems, but it does produce some of the knowledge needed to solve some of them.) In short, optimism in the capacity of research is well founded. (Whether research will continue to be conducted is another question to be addressed in Ch. 13, Section 4.) Without such rational faith in research no long term research project would ever be undertaken. Radical skeptics and dogmatists have no such faith and as a consequence they never discover or invent anything but errors in others.

Our faith in inquiry, though great, is not boundless: it is tempered by the following facts. First, many solutions that seemed flawless at the time they were obtained turn out to be defective in some way or other (even in pure mathematics); consequently they have to be repaired or even abandoned. Second, some problems are given up before they are solved, either because they fall out of fashion, or because they are intractable with the available means, or because they are seen to be unimportant, or because of new unfavorable cultural or political circumstances. Third, there are limits to the information we can obtain and to the number of problems we are capable of tackling: we are not omniscient and we do not want to know everything that can be known. (Information overload can be as bad for original research as lack of information, for it is confusing and inhibiting. And lack of information is sometimes unavoidable, particularly regarding past events: their traces have often been obliterated by other events, so there is no record to be studied.) Fourth, there are certain problems that, because of their breadth and depth, are likely to stay with us forever: they are the scientifico-philosophical and technico-philosophical problems.

Think of "What is matter?", "What is spacetime?", "What is life?", "What is mind?", "How can we know?", "What can we know?", "What is good?", "What can we do and what ought we to do?", and "What does the good society look like?". These are not mere puzzles, let alone illegitimate philosophical questions, as the pseudoanalytic counter-revolution in philosophy has labelled them. (Cf. Warnock, 1958.) Nor do they arise from a misuse of words, so they will not be solved, let alone dissolved, by linguistic analysis—the way Wittgenstein and his disciples would have us deal with philosophical problems. (Remember the famous pronouncements: "philosophical problems arise when language *goes on holiday*"—Wittgenstein, 1953, p. 19—"the clarity we are aiming at is indeed *complete* clarity. But this simply means that the philosophical problems should *completely* disappear"—*ibid.*, p. 51—and "The philosopher's treatment of a question is like the treatment of an illness"—*ibid.*, p. 91.)

Those wide scope problems are the very *raison d'être* of entire disciplines, and they are unlikely to receive final solutions even though they are being investigated and progress is being made in solving them. All grand problems are like hydras: as one head is cut another sprouts. But, unlike Hercules' hydra, they will continue to defy mankind as long as there are curious people left. Progress changes the problematics without shrinking it. On the contrary, the more we get to know, the more and tougher problems can we formulate. Progress in any epistemic field is measured not only by the problems that get solved but also by the unsolved problems that are being investigated.

As long as there are rational beings they will face and seek problems of various kinds: (*a*) *practical* problems affecting individuals and social systems; (*b*) *cognitive* problems of either the substantive or the methodological kind—or, in a different classification, empirical or theoretical; and (*c*) problems of *valuation and morals*. A correct, even if only approximate, solution to problems of any such kinds requires some knowledge, sometimes available, mostly to be sought. Moreover, the investigation of any problem of some kind poses further problems of different kinds. Thus practical and moral problems require some substantive knowledge, and the investigation of cognitive problems requires abiding by certain methodological and moral norms. In short, the composition of each problem system is heterogeneous. The heterogeneity is sometimes such that it calls for a multidisciplinary or even an interdisciplinary approach.

5. CONCLUDING REMARKS

What distinguishes modern thought from that of our ancestors 50,000 years ago is not so much better brains as the kind of problems some of us tackle and the way we approach them. Whereas our remote ancestors presumably devoted their whole brain power to solving day by day survival problems, agriculture and village life gave some people, from about 10,000 B.C. on, enough leisure to apply their wits to speculative problems, as well as to try and solve practical problems more rationally than by blind trial and error. That social revolution—the beginnings of agriculture and urbanization—caused a shift in problematics and in methodics, and this in turn gave us the type of knowledge we are proud (and sometimes ashamed) of. And, of course, some of the outcomes of such intellectual endeavors have in turn elicited profound changes in our life styles—not always for the better.

Problems are the root and the fruit of original research, be it in science, technology, or the humanities. Therefore we should correct the neglect of the very concept of a problem in the philosophy of science. More research on the logic, semantics, and heuristics of problems should be able to help us work better at finding and solving problems. In particular, it should help us avoid barking up the wrong tree and devoting our talent to pseudoproblems (such as trying to reduce social science to psychology or to biology) and *cul de sac* problems, i.e. miniproblems whose solution would make no noticeable difference to the state of knowledge or of society.

The specialist may be trusted to spot wrong solutions, but spotting wrong problems takes more than professional competence, particularly when an entire army of knights has been tilting at windmills for some time. In such cases there is a (bad) model to be imitated and even a vested interest in persisting in the effort. (The quixotic attempts to build logics of analogy and of induction, to reduce all philosophical notions to those of probability or of information, and the exact but barren investigations in subjective probability, modal logics, and possible worlds, come to mind.) As Ackoff (1974, p. 8) puts it wisely, "We fail more often because we solve the wrong problem than because we get the wrong solution to the right problem".

A touch of epistemology might also help spot and avoid trivial problems, i.e. problems which, though well formulated, are unlikely to add significantly to human knowledge. All sciences, particularly though not exclusively the less advanced ones, are littered with trivial research, the only functions of which are to allow some individuals to make a living, and to render the volume of the literature unmanageable. Think of the human and material resources that have gone into confirming the well-known platitudes that talent depends on both genetic factors and opportunities, that groups composed of competent people perform better than bunches of incompetent people, that persons with low self-esteem are reluctant to participate in group work, that group discussions are dominated by high ability or highly vocal individuals, and so on. No wonder the announcement of such trivialities is usually greeted with the sneering question 'So what?', or (in Newyorkese) 'So what else is noo?'.

In conclusion, valuable inquiry is triggered by good problems. How does one get new problems? By reflecting critically on old problems and their solutions, as well as on new situations. And how does one get ideas to solve the new problems? By struggling with them. Problems, in sum, are the very heart of inquiry.

CHAPTER 8

CONJECTURING

Solving problems, whether practical, cognitive or moral, involves conjecturing: that this fruit is edible and that animal dangerous; that these rains announce a cold spell, and that stranger is friendly; that this stone may be suitable for fashioning an axe, and those twigs for kindling a fire; that this action may benefit my kin and that one hurt them—and so on and so forth. Every problem demands making some conjecture or other, and every conjecture poses the further problem of finding out whether it is adequate (true, efficient, or good).

We seldom face experience without some preconception and anticipation, and we never undertake scientific experimentation or technological design without some hypotheses. It is not that we are perfectly rational beings, nor that we abstain from perceiving or doing until we have formed a true theory or a suitable plan, nor that every datum is preceded by some careful cognition, but that perceiving, thinking and doing go often hand in hand. To be sure we are often taken by surprise, unprepared to react to or understand what happens to us. But only those who know and expect something can be surprised, and only those who make some conjectures about surprising events, and care to check such conjectures, can hope to survive, or at least to subsist as rational beings.

1. FROM PRECONCEPTION TO HYPOTHESIS

1.1. *Psychobiology of Conjecturing*

We all know that the animals capable of learning can learn to expect certain events to follow others, and to perform certain actions after certain events occur, or in order to attain certain goals. But how do they learn: by reinforcement of the adequate expectations and actions, or by forming and trying out alternative hypotheses? Behaviorism gives the former answer, and cognitivism the latter. (Note, by the way, that this disjunctive problem contains the hypotheses we wish to discuss.) Perhaps both schools are right: maybe the two learning mechanisms operate in the higher animals, whereas the lower ones learn only, if at all, by reinforcement. But, since this chapter happens to be about the nature of hypotheses and their role in problem solving, in this case we are likely to learn more from cognitive psychology than from behaviorism.

Classical—i.e. mentalistic and introspective—psychology had paid much attention to thinking and, in particular, to hypothesis formation, though usually without employing the experimental method. (See Claparède (1934) for an excellent detailed review.) In particular, Gestalt psychology had emphasized the role of hypothesis in problem solving among primates, but it had ignored the question of testing. The animal faced with a problem was said to start by making a few haphazard trials, then engage in serious thinking until struck by a clap of insight, upon which he acted in the correct way. There is some truth in this picture but it is incomplete and inaccurate. Preliminary exploration is often systematic (because guided by experience, by cues or by clues) rather than random. And hypotheses are tested (in imagination or empirically) no sooner they are formed. Moreover the outcome of such tests often suggests new hypotheses or poses new problems.

The breakthrough in the experimental study of the role of hypothesis in problem solving was inspired by Lashley (1929) who, in an influential book, suggested that learning proceeds by making hypotheses and putting them to the test. Krechevsky (1932) and Tolman and Krechevsky (1933) took up this idea and tested it in rats. They found that the rat does not behave haphazardly during the early stages of a problem-solving process but proceeds quite methodically, "attempting various solutions and giving them up when they fail, until he hits finally upon the 'correct' one" (Krechevsky, 1932). This work was severely attacked by the behaviorists, in particular K. Spence, and so the incipient American branch of cognitive psychology went underground. (It never ceased to be in the foreground of European psychology.)

The hypothesis that the higher vertebrates learn by making hypotheses and testing them resurfaced in America with *A Study of Thinking* by Bruner, Goodnow and Austin (1956). It is now being investigated vigorously both in the laboratory and mathematically. (Cf. Restle, 1962; Levine, 1966, 1974; Mayer, 1977.) One finding of this research is that human learners are not passive: they do not respond always in the same case to a given stimulus, but their response depends upon their experience, expectations, etc. Another is that the correct response is not strengthened gradually by reinforcement: in some learning tasks the subject can learn at a single trial. Both results, known or suspected by nonbehaviorist psychology, refuted behaviorism.

Restle (1962) has interpreted these experiments in the following way. At the outset of a learning task (discrimination learning, maze learning, or concept formation) the subject has a universe of hypotheses or strategies. He samples them first at random, and acts accordingly. If he hits on the correct hypothesis or strategy, he sticks to it without making any further errors. Otherwise he tries again by resampling. So he may choose the just rejected hypothesis with the same probability as any other one—a rather implausible assumption. (The only rationale for it seems to be that it allows its author to model the process on that of a lottery with replacement, and so with constant probability for each strategy or hypothesis.)

This model seems unsatisfactory if only because it contains the zero-memory hypothesis just commented upon. But there is more. If a problem is soluble at all, then either there are methods for tackling it or there are none. If there are, then the subject may know some of them or he may not know any. If he does, then he is likely to try first the simplest, or the one he remembers best. And if he does not know of any such method he may give up or give it a try. If the latter, he is likely to proceed by groping (*tâtonnement*) but not wholly in the dark: one is always guided by certain clues, and one can always exert one's imagination.

That hypothesis formation is seldom haphazard is suggested by observations and experiments on the way children and adults solve qualitative combinatorial problems. Take for instance anagrams, and in particular the 5 letter word 'ilrta', meaningless in English. There are 119 variations of this string of letters but only a few of them, such as 'trail' and 'trial', belong to the English vocabulary and therefore qualify as solutions. Experiment shows that, typically, subjects do not try all the 119 permutations but begin with the combinations that have the highest frequencies, e.g. 'tr' rather 'rt'. (The strength of such clues is such, that subjects find it usually difficult to break the letter pairs that occur frequently, so a problem word such as 'beach' is harder to crack than its anagram 'hbcae'.) In any case, the time it takes most subjects to solve such problems varies typically between 10 seconds and 2 minutes, and is therefore too short to allow for the examination of all the alternatives. (See, e.g. Mayer, 1977.)

So, even in the absence of methods one usually employs some clues, not always explicit, i.e. knowledge—explicit or tacit—of certain regularities that he has learned before. That is, hypotheses, even when they seem to come by sudden illumination, do not jump out of nothing but are somehow generated from available knowledge. Exactly how they are formed we do not know yet, but it is likely that thinking up a new hypothesis is identical with the activation of a new psychon. It would also seem that there are several different modes of generation. Let us consider quickly a few of them.

To begin with there is induction from observed cases, either from very many or from just a few or even one. Thus it usually takes a young dog a single disagreeable experience with fire to form the generalization that all fires burn. Because of the universality of induction among the higher vertebrates, it cannot be the mark of science—*pace* the opinion of most inductivist philosophers. What is a (though not the) mark of science is the control or test of generalizations, inductive or not, by means of observation or experiment, as well as the attempt to explain them—e.g. to understand why fires burn. Without such tests, i.e. without going beyond the empirical sources of our hypotheses, we are likely to accumulate false beliefs, such as the time-honored ones that every existent fulfils some function in a well designed order; that all celestial bodies revolve around our planet; that every comet announces some major catastrophe; that all societies are stratified; that all conquered or colonized peoples are inferior; that all humans believe in deities; and that all knowledge comes from the senses. After all, every one of these superstitions has a good empirical basis.

At other times we form hypotheses by noting associations. Thus if we know that A and B often occur together, or one after the other, we are likely to conjecture that A and B are somehow correlated or even functionally related. (This requires the prior nontrivial isolation of A and B from the jumble of phenomena, or even the hypothesizing of their existence.) Unless we are engaged in scientific or technological research, we are likely to care only for the joint (or successive) occurrence of A and B, disregarding the three remaining cases, which are just as important, namely A and not-B, not-A and B, and not-A and not-B. For example, even experienced nurses will "see" nonexisting correlations between disease and symptom for disregarding such negative cases; and, in general, "adult subjects with no statistical training apparently have no adequate conception of correlation" (Smedslund, 1963, p. 165). The fallacy of nonexisting association is of course at the root of many superstitions, such as the beliefs in magic, astrology, and miraculous cures.

At other times we form hypotheses on the strength of similarities, real or imaginary. Thus Faraday invented the concept of electromagnetic field by analogy with the concept of an elastic body; and Darwin drew inspiration from his readings about economic competition and overpopulation. At still other times we form hypotheses with the help of general principles. Thus if we note that the total energy of a system has decreased or increased, we use the principle of conservation of energy to conjecture that the system is not closed, so that there has been an energy leak.

However, the deepest and therefore most important hypotheses are not suggested by recurrent phenomena, repeated associations, or real analogies, but they are newly invented. The reason is that, by definition, a deep hypothesis concerns facts that are not accessible to direct observation. The phenomenal world is complex and messy. A meticulous recording of everything we perceive is unilluminating because of excessive detail. To understand what we perceive we must imagine imperceptible things or connections, or analyze in imagination what perception presents as a solid block.

Thus Harvey, who had at his fingertips just as many anatomical data as his predecessors, was the first to conjecture that the heart, the arteries and the veins form a system of closed circuits. Part of this conjecture was the hypothesis that there are capillaries, invisible to the naked eye, that join the ends of the arteries to the beginnings of the veins. This hypothesis was confirmed, with the help of the newly invented microscope, only after Harvey's death. Likewise Maxwell, guided by the general hypothesis that all electric currents are closed, postulated the existence of a displacement current between the plates of a discharging condenser. This hypothesis was confirmed after his death. And Ramón y Cajal, using Golgi's staining technique, showed that every region of the brain is composed of neurons. Although his preparations exhibited only discrete cells, he conjectured that the brain forms a system, and this hypothesis was confirmed only half a century later with the discovery of synaptic transmission. In all three cases the initial data had been insufficient even to describe what happened. In all three cases the hypotheses turned out not only to be true but also tremendously fertile. But although deep, audacious and path-breaking, all three conjectures were initially just explicit and testable educated guesses. In this regard they were no different from the thousands of hypotheses, of varying grades, that every active scientist, technologist, humanist or professional frames in the course of a year.

1.2. *Hypothesis*

In mathematics the word 'hypothesis' designates an assumption such as a datum of a problem. For example, in proving a theorem about right angle triangles we may, at a certain point, need to remind ourselves that, *by* *hypothesis*, the triangles in question have a right angle. Tentative statements, such as candidates to theorems, are called 'conjectures' in mathematics. We shall be concerned throughout with conjectures.

We make guesses or conjectures all the time, in ordinary life as well as in the various sciences, technologies, arts, and professional activities. However, not all of our conjectures deserve being called 'hypotheses'. Only the *educated guesses* that are *formulated explicitly* and are *in need of testing* qualify as hypotheses. The formulation must be explicit but need not be propositional: the design of an artifact, a historical reconstruction, the reconstruction of a lost original, and the program of a course of action qualify as hypotheses. See Figure 8.1.

We have employed the expression 'educated guess' to designate a guess that, far from being wild, is compatible with some background knowledge. For example, suppose we are asked to guess the number of children in Peanutia, a Third World country with one million people. A wild guess is an arbitrary number between 0 and 1,000,000. An educated guess is this: since in most Third World countries children constitute half the population, Peanutia has about 500,000 children. This guess is educated and it can be put to the test by looking up the latest census figures: it is a testable hypothesis. The guesses of all experts are of the same kind and are formed in the same way: by combining bits of specific information (e.g. the total population of a certain Third World country) with generalizations (e.g. that in Third World nations persons up to 15 years old constitute half the population).

Some of our hypotheses concern observed or observable facts, like the conjectures we make while driving, about the congestion of traffic ahead, or the coming rainfall, or the possible outcome of a test (e.g. an experiment), or the performance of an as yet untried artifact. Other conjectures are about unobserved and perhaps unobservable facts, like those about other people's intentions we make while driving, or the behavior of our relatives while they are not being observed, or the efficiency of a new form of organization, or the nonphenomenal properties of things of any kinds.

The hypotheses concerning as yet unobserved entities or properties must be conceived, of course, before the corresponding direct or indirect observations can be performed. At the time of conceiving such hypotheses there may be data, but all they do is to pose the problem of explaining them and constraining the possible hypotheses. Moreover the hypotheses are needed in order to design observations aiming at showing the existence or nonexistence of such entities or properties. Recall some outstanding

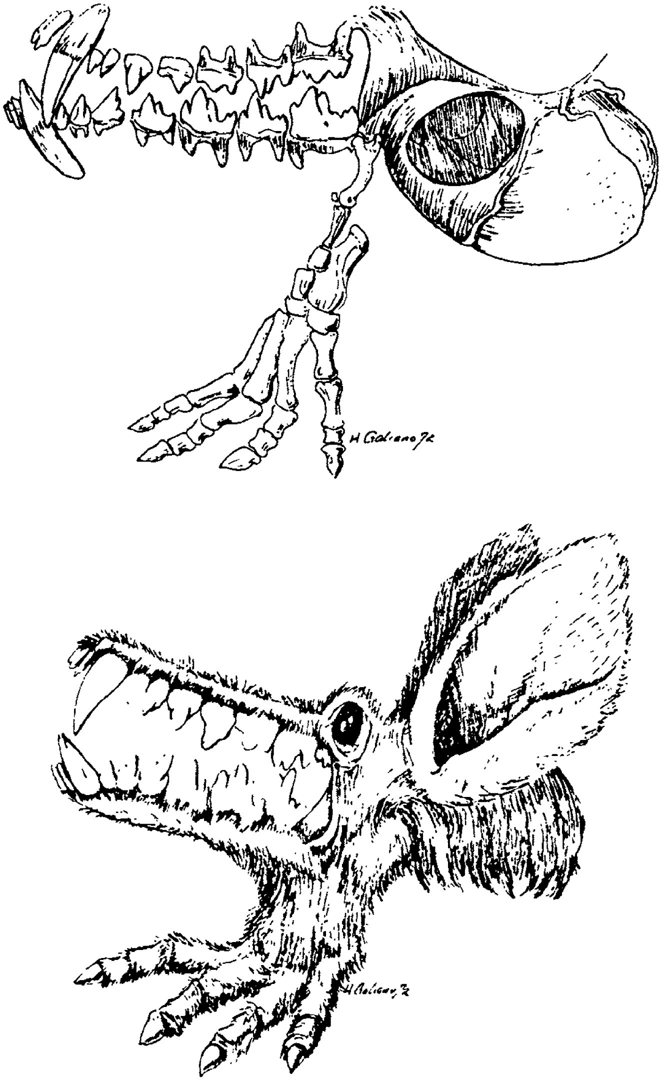


Fig. 8.1. Hypothetical reconstruction of the past: the case of fossil mammals according to McKenna (1976, pp. 246–7). (*a*) Strict adherence to the data (fossil bones) would lead us to believe, as the late A.S. Romer used to say in jest, that the extinct mammals were merely teeth that lived, gave birth directly to more teeth, and died. (*b*) We know better: the early mammals had also a foot, an auditory apparatus, and fur.

examples of this kind of hypothesis: atoms were hypothesized twenty-three centuries before their existence could be established experimentally; genes and antibodies were postulated half a century before being observed and identified chemically; nuclear, interatomic and intermolecular forces are hypothesized in order to explain the outcome of scattering experiments, the main function of which is to furnish some indications about those forces.

Some of the unobserved facts we make hypotheses about are in principle observable and, with ingenuity and luck, are eventually observed. Others are unobservable *in principle*—at least according to our present knowledge. A clear case in point is that of historical hypotheses, i.e. hypotheses about the past of the universe or of any part of it. All such hypotheses are about alleged facts that, not existing any longer, are unobservable in principle. To be sure, some such facts are of the repetitive kind and so we can use our knowledge about current facts of the same kind. For example, the emission and absorption of light can be assumed to "obey" now the same laws it "obeyed" a thousand million years ago; and we can assume that the earliest hominids had most if not all of our own basic needs.

But other facts are unique or unrepeatable. Hence if past and unobserved they must be reconstructed hypothetically on the strength of whatever traces they may have left, and of some regularities. For example, the feeding habits of animals belonging to extinct species can be reconstructed in either of the following ways. One is to analyze the undigested remains of the excreta, which are rather exceptional findings. If this method cannot be used for lack of data, the palaeontologist attempts to reconstruct the physiology of the animal on the basis of its fossil remains—an exciting and chancy exercise. A third, no less fallible way, is to extrapolate back from the known feeding habits of living relatives. The best procedure would of course be to combine all three, but this is seldom possible. And in any event each of the three methods involves making hypotheses. In general, the shorter on empirical evidence the longer on speculation, the longer on evidence the longer on well-founded hypothesis, and the longer on the latter the longer on new discoveries or inventions. To abstain from speculating altogether is to condemn oneself to living in the present without understanding it. Not speculation in itself but only wild, i.e. groundless and untestable, speculation is to be shunned. Let sound speculation flourish.

Every inquiry, even the simplest data gathering operation, involves some hypotheses about the object of study as well as on the manner of studying it. (The former are substantive, the latter methodological hypotheses.) For example, in going about their data gathering, anthropologists and field linguists have usually assumed that the uneducated elders of the tribe or the village were the repositories of tradition and language in its purest form, before being corrupted by the radio, tourists, or visiting scientists. In this way they collected many data that were not representative of the population as a whole and even, in some cases, histories deliberately made up to please or fool the scientists. That such mishaps are difficult to avoid is beside the point. The point is that data, whether in science, technology, the humanities, or professional life, are always sought in the light of some substantive and methodological hypotheses: before rushing to make observations or experiments we must choose the kind of data we want and must decide how best to obtain them. What to do with the data once collected is another story.

According to empiricism the task of the researcher is finished once he has got his data. At most he may summarize and cautiously generalize the information thus obtained, i.e. he may formulate low level hypotheses. (A hypothesis may be said to be of the *low level* kind if it contains no substantive or technical concepts other than those occurring in the data it covers. Data-fitting curves are of this kind.) No doubt such a task is to be performed, if only because the low level hypotheses are raw material for the theorist. See Figure 8.2. But some of us want not just to know what is the case but also to understand, and this calls for inventing further hypotheses—explanatory hypotheses. In short, the process is not from data to hypotheses (the positivist schema) but from problem to hypotheses, and from these either to data (in the case of empirical research) or to further hypotheses (in the case of theoretical research), and from either to control or test—and then back to the original problem or forward to yet another problem, perhaps one generated by either the hypothesis itself or its test.

Think of any of the research processes that culminated in some of the most interesting findings of our century—e.g. the quantum theory of atoms, which is not a data package but a high level construct containing no

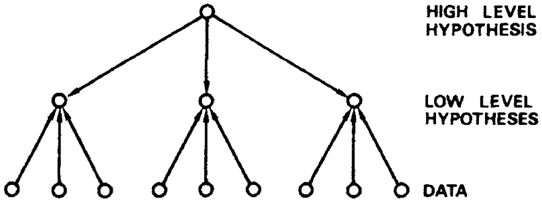


Fig. 8.2. High level hypotheses colligate and entail low level hypotheses, which in turn summarize and generalize (but do not entail) data.

data proper. The motivation for building it was the failure of classical physics to explain some facts known of old as well as some newly discovered facts. There was nothing to induce from: induction came only at the end, when evaluating the theory in the light of experimental results. Or think of the many hypotheses involved in any model of the economy of a whole nation, such as an econometric model of the US economy. Or of the many hypotheses involved in any long term development plan—or, for that matter, in any plan to turn back the social and economic clock. Or, finally, think of the history of television and the host of hypotheses of all levels, including some taken from electrodynamics and others from electron physics, involved in it.

Are there any rules or methods for conceiving hypotheses? A number of outstanding thinkers, among them Leibniz, have claimed that they had invented, or could invent, an *ars inveniendi* or logic of invention and discovery. The idea was to invent a handful of precise and comparatively simple recipes wherewith anybody could discover truths regardless of his background, talent, and motivation. Of course nobody did in fact exhibit any such universal rules, and the project died out. A few optimistic philosophers (notably Hanson, 1958) tried to revive it in our century but, predictably, they too failed. The project of inventing an *ars inveniendi* is chimaerical because originality consists in breaking new paths, not in following existing ones. It would be far less wrong-headed to investigate which chemicals, and at what concentrations, favor or inhibit original conjecturing. And it is a legitimate problem for sociologists and historians to find out what types of social condition stimulate or inhibit conjecturing. (Cultural freedom, though necessary, is insufficient. Affluence, though helpful, is neither necessary nor sufficient. A tradition of learning is ambivalent. And a philosophy contemptuous of hypotheses is inimical.)

However, there are a few methods for generating hypotheses from data. Any interpolation formula, allowing one to construct a continuous curve (a polynomial) out of a finite number of data, is such a technique. The technique for calculating linear regressions is another. These and other techniques can be incorporated into computer programs. In fact computers can be programmed to generate hypotheses of the following kinds: "A is associated with B", "A increases with B", "A decreases with B", and "C makes A and B correlated" (Hájek and Hávranek, 1978). More advanced programs, particularly DENDRAL, are employed in organic chemistry to form the plausible structure diagrams of molecules on the basis of their composition and mass spectra, the valence of their component atoms, and a few other data and hypotheses (Lederberg and Feigenbaum, 1968). In short, it is possible to mechanize induction.

All such algorithms are of great practical value to process huge masses of data, as is the case in statistics, chemistry, medicine, and some other fields. However, it must be kept in mind that the resulting hypotheses are all *low level* ones and therefore of no great theoretical interest. In fact the hypotheses generated by algorithms out of data (*a*) colligate data (to use the term introduced by Whewell (1847)); (*b*) involve only substantive concepts occurring in the data (e.g. those of atom and valence in chemistry), and (*c*) are mathematically extremely simple-minded.

The generation of powerful hypotheses, such as differential and integral equations, is beyond the reach of any algorithm or machine program. Even simple law statements such as Snell's law of refraction cannot be generated from a bunch of data. What one can do with a computer is to *check* whether a given high level hypothesis fits a certain data base. In sum, computers can help process data and generate low level hypotheses, as well as find out how well or poorly the latter fit the data. But they cannot invent any high level hypotheses because the latter involve, by definition, concepts that fail to occur in the data.

We do not know of any rules or methods for generating high level hypotheses such as the hypothesis of natural selection, economic determinism, or the Schrödinger equation. We just grope for high level hypotheses, helped or hindered by induction or analogy, by general principles or mathematical forms, and always spurred by imagination, a knack for discerning or guessing patterns, and the philosophical conviction that there *are* patterns to be uncovered. All of the high level hypotheses are invented in an unruly way and perhaps also unconsciously. Only their precise formulation and checking must be methodical and fully conscious.

Philosophers, in sum, are no longer interested in the chimaerical project of setting up a logic of discovery or of invention. They work, on the other hand, on such problems as: (*a*) defining the very concept of a hypothesis, (*b*) classing hypotheses (e.g. into empirical and theoretical, or low level and high level, substantive and methodological, etc.); (*c*) unearthing the clues that led or misled to the formation of some hypotheses of historical interest; (*d*) establishing the conditions a hypothesis should fulfil in order to deserve being examined theoretically or empirically; (*e*) establishing the conditions for concluding that a hypothesis has been satisfactorily confirmed. These are genuine problems, and open ones for the most part. (Because of the empiricist tradition in epistemology they are often lumped under the misnomer *Induction.*) We shall address some of them in this book.

2. SCOPE AND DEPTH

2.1. *Scope*

Hypotheses come in all scopes or extensions, and the real scope of a hypothesis is not always apparent but must be dug out by logical analysis. We distinguish the following kinds of hypothesis with regard to scope:

1. *Singular* hypotheses, such as "This machine should work", and "That social program should decrease unemployment".

2. *Particular*
   2.1. *Indefinite* particular hypotheses, such as "There are black holes", which specify neither place nor time and are therefore hard to refute.
   2.2. *Definite* particular hypotheses, like "There are decision-making modules in the frontal lobes of all normal higher vertebrates".

3. *General*
   3.1. *Bounded universal* hypotheses, such as "All Renaissance intellectuals were generalists", that are restricted to certain places or times.
   3.2. *Unbounded universal* hypotheses, such as the laws of physics.

Scope is no ambiguous indicator of importance. Thus the hypothesis that such and such conflict was a trade war may be more important to us, who live in the shade of world wars, than many a generalization of mirmicology or of malacology. Besides, some hypotheses are singular in one respect but universal in others. For example, the astronomical and geophysical laws concerning our planet are referentially singular but they are universal with respect to time. For example, "The Earth spins" abbreviates "For all times since its formation the Earth spins". Likewise every statistical correlation, e.g. between the height and weight of persons, is a universal statement. Indeed, "Variables A and B are strongly correlated" means that, for every value of A, the corresponding value(s) of B lie close to the linear regression line.

Nor is scope, or quantifier type, a precise indicator of degree of generality: logic is too coarse a tool of analysis as compared with mathematics. For example, "For all $x, f(x) = 0$" is a special case of "For all $x, f(x) = a$", which in turn is subsumed under "For all $x, f(x) = a + bx$", and so on. Likewise rest is a special case of motion, uniform motion (or stationary state) a more general one, and uniform acceleration still another, even more general kind. Neither of the above generalizations is inductive, for neither consists in jumping from singulars to universals. They are generalizations from species to genera, from these to families, and so on. (Recall Ch. 6, Section 2.2.) Moreover, as a matter of historical record some of the higher order generalities were conjectured before the special cases were investigated.

Empiricists have always cautioned against the sin of overgeneralization, and have advised us to stick to singulars, the only genuine existents. There is no need for such warnings, since empirical tests will correct any excesses. Tests will also correct the dual mistake of undergeneralization, by showing that what was initially thought to hold only for a given set of facts holds also for a superset of it. (A notable recent case was quantum mechanics, originally conceived for atoms only, but later generalized to molecules, atomic nuclei, and even certain macrosystems such as superconductors.) Of the two complementary mistakes, undergeneralization is more dangerous than its dual because it may consist in not making any generalizations at all, or in stopping at low level generalizations. After all, a high level general hypothesis is better than none, for it stimulates thinking and experimenting—and it is fun.

Generality is related to inferential power, or the capacity a proposition may have of generating further propositions with the help of some formal tools. Actually even the most modest singular proposition, such as a datum, may be said to have an infinite deductive power, for it entails infinitely many conditionals where it occurs as a consequent. (I.e., e entails $h \Rightarrow e$, with h arbitrary. In fact, suppose the inference invalid, i.e. $h \Rightarrow e$ false. This holds only if h is true and e false—which contradicts the hypothesis that e is true.) Such profligacy of singulars is uselesses because, by entailing anything, they generate nothing in particular. (Inductivists, take note.) Genuine deductive power consists in the possibility of specification, in particular instantiation.

We distinguish two genera of hypothesis with regard to specifiability: the unspecifiable and the specifiable ones. Singular propositions, such as "b is an F", and particular propositions, such as "Some A's are B's" (or even "Most A's are B's"), do not entail anything definite. Likewise the outcomes of statistical analyses, such as "f percent of A's are B's", and "The standard deviation of the F distribution is such and such", entail nothing. In general statistical statements, unlike general probability statements (e.g. on probability distributions), are deductively barren. They are terminals of inquiry and, at most, they pose the problem of setting up models for explaining them.

We distinguish two species of specifiable hypotheses: immediately and conditionally specifiable. An *immediately specifiable hypothesis* is one from which singular hypotheses can be derived by assigning definite values to variables. The empirical generalizations and the lowest level theorems of factual theories comply with this condition. For example, if a function f solves a certain ordinary differential equation, it follows that its value at point b is $f(b)$. (That this value may be difficult to compute is beside the point.) A more interesting example: the basic equations of a fundamental theory, such as classical dynamics or classical electromagnetism, are specifiable to species and to individuals by specifying the functions and parameters occurring in them—e.g. by specifying the force function, the mass value and the initial conditions in Newton's law of motion, or the four-current and the boundary values in Maxwell's laws.

A *conditionally specifiable hypothesis* is one that can be "applied" (made to refer to) individual cases, or subordinate species, upon suitable formal or semantical transformations. For instance, differential and integral equations must be solved and assigned factual interpretations before they can be so "applied". Mathematics takes care of the indicated formal transformation, but it is the responsibility of the scientist or technologist to "read" the resulting formulas in terms of concrete things, their properties, and changes in the latter. That is, the mathematical formula must be adjoined meaning assumptions or correspondence rules. (Cf. Vol. 2, Ch. 7.)

Some hypotheses are so extremely general that, in conjunction with particular or even singular bits of information, they entail further general—though somewhat less general—hypotheses. For example, if a given organic molecule consists of subsystems A and B, it can be hypothesized that A and B are the precursors of the molecule (i.e. have preexisted it), on the strength of the general postulate that every complex system has arisen by the assembly of its components. This is so general a hypothesis that it belongs in the intersection of science, technology, and philosophy (Vol. 4, Ch. 6).

Scientific, technological and humanistic research are conducted with the help or hindrance of a number of philosophical hypotheses. We mentioned a few such hypotheses of an epistemological kind in the Introduction, Section 6, as well as some hypotheses of the ontological kind in Vol. 3, Introduction. These philosophical hypotheses do not occur among the findings of scientific, technological or humanistic research, but they guide research and are controlled by it. One such hypothesis is that logic and mathematics have no factual content and therefore (*a*) can be used in any epistemic field and (*b*) are not testable by any empirical procedures—i.e. they are a priori. Another is that we can get to know more than appearances, i.e. that it is legitimate, nay mandatory, to frame and use transempirical concepts and hypotheses. A third, that the object of research is distinguishable from the knowing subject. (To be sure, the latter may influence the former, but no such influence would be possible if they were indistinguishable, i.e. one rather than two.)

Every one of these and other philosophical hypotheses is presupposed by modern research, but this does not entail that every epistemic field has a permanent philosophical basis, or that philosophy can be the ultimate arbiter. What it does entail is that there is no separation between philosophy and the other epistemic fields; consequently that neither of them can advance without the other's support and criticism. Philosophers are not equipped to approve or question any specific hypotheses in fields other than their own—unless of course they happen to be conversant with them—but they are certainly competent to approve or question the philosophical hypotheses presupposed by such specific hypotheses. Although such philosophical presuppositions of research provide no ultimate foundation, they do constitute a provisional basis to be further analyzed, systematized, supported, or corrected. They are true to the extent that they facilitate the search for truth, barren if they neither lead nor mislead, and false if they hinder such search in any way.

2.2. *Depth*

Scope is only one of the dimensions that determine the size of a hypothesis: the other is depth. We stipulate to call a hypothesis *deep* only when it contains transempirical concepts, and otherwise *superficial*. For example, the hypothesis that syphilis is only a skin disease is literally skin-deep. It was held for thousands of years and it sheltered a number of fantastic conjectures about the origin of syphilis—e.g. that it was ordained by the stars, and that it was a divine punishment for the sin of fornication. In the wake of the microbiological revolution initiated by Pasteur, the programmatic hypothesis was formulated that syphilis was caused by a microbe. Methodical investigation into this deep (transempirical) hypothesis proved successful: in 1905 *Spirocheta pallida* was discovered and incriminated. (Cf. Fleck, 1935.) In turn, this finding posed the technological problem of getting rid of the microbe—a problem solved only with the advent of antibiotics half a century later.

Something similar is happening nowadays with regard to cancer. Surgical practice and observation have produced plenty of accurate descriptions of the outward appearance and location of cancerous tumours; clinical observation has yielded many data on the symptoms of cancer; and histological observation has completed the picture. But neither of these procedures could possibly tell us what the mechanism of cell proliferation can be, and consequently neither of them can give us a clue to the prevention or cure of cancer. On the other hand cell biologists, molecular biologists and virologists are on the right track and are learning more and more about such mechanism. The ultimate goal is a detailed and precise (if possible mathematical) model of cell proliferation—an explanatory, not just a descriptive one. But we will not get such a system of deep hypotheses if we continue to believe that knowledge can be gained only through observation or practice; if researchers continue to adhere to the empiricist epistemology, which shuns transempirical concepts and therefore deep hypotheses; and if most science administrators continue to look upon the problem as one in medicine rather than in biology.

According to our definition, a hypothesis qualifies as deep if it contains at least one transempirical concept—aside of course from the logical or mathematical concepts occurring in it. Now, we must reckon with several degrees of depth. Thus a black box model of a system, i.e. one that ignores its internal composition and structure, is superficial, for it represents only the observable inputs and outputs. If we allow for internal states of the box, i.e. build a grey box, our representation of the system becomes somewhat deeper. Still, this depth is not impressive as long as the internal states are not identified as states of the internal mechanism of the box.

For example, the simplest model of a system is the vending machine with just two internal states, the ground state and the excited state: Figure 8.3. In the absence of external stimuli the machine stays in the ground state, i.e. it lacks initiative. (When it does exhibit initiative and starts delivering goods without being stimulated, it is judged to be out of kilter.) When a stimulus of a certain kind, e.g. a coin, impinges on the machine, the latter jumps to the excited state, produces a response of a certain kind (e.g. delivers a candy bar), and returns to its ground state. Such a system can be realized in many ways, i.e. it is compatible with a whole family of internal mechanisms—mechanical, electrical, chemical, etc.

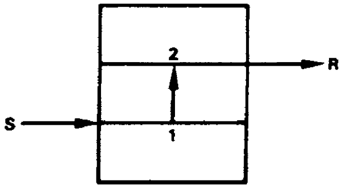


Fig. 8.3. A two state grey box model of a system. When stimulus S impinges on the system, the latter jumps from state 1 to state 2, produces response R, and then decays back to state 1. If the machine is probabilistic, it responds to the stimulus with a given probability p, so it steals dimes a fraction $1-p$ of the times it is stimulated. A simple mathematical representation of such probabilistic machine is the matrix equation $R = MS$, where $M = \begin{Vmatrix} 0 & p \\ 1 & 0 \end{Vmatrix}$ and $S = \begin{Vmatrix} 1 \\ 0 \end{Vmatrix}$, whence $R = \begin{Vmatrix} p \\ 0 \end{Vmatrix}$, with $0 < p \leq 1$.

The consumer may not mind the kind of mechanism as long as it is functional. But of course scientists and engineers are keen on discovering or designing the mechanisms of the systems they study, for they wish to understand or master them. In particular the neuropsychologist wishes to understand the neural mechanisms of cognition: he cannot remain satisfied with the "cognitivist" or "functionalist" dogma that only the "software" or program matters, the "hardware" or neural stuff being irrelevant. If a philosopher insists on remaining ignorant of such mechanism, it is his privilege. But he has no right to press for a commitment to such ignorance as the right strategy for investigating the mind. (More in Vol. 6.)

The search for, and description of, mechanisms involves *mechanismic* or *dynamical* hypotheses, i.e. deep conjectures about mechanisms of some kind—mechanical or electrical, chemical or cellular, organismic or ecological, economic or cultural. The hypotheses that gas pressure is due to molecular impacts, and that extreme class differences cause discontent, are well known examples of mechanismic conjectures. Nonmechanismic hypotheses, whether or not they contain transempirical concepts, are called *phenomenological*, *kinematical*, or *black box* conjectures. The thermodynamical relations between pressure, volume, and temperature, and the stimulus-response hypotheses of behavioristic learning theory, are typical examples of phenomenological conjectures.

Kinematical or black box hypotheses are easier to frame and test than their dynamical rivals. Moreover kinematical hypotheses contain adjustable parameters and can therefore be made to fit the data more perfectly than the corresponding dynamical hypotheses. It is always easier to describe a state or a change of state than to explain how it comes about. However, we need hypotheses of both kinds for, after all, a description of facts must be available before we can proceed to explain it. Thus phenomenological hypotheses, even if they fit accurately the known data, pose the task of framing mechanismic hypotheses capable of explaining the former. We must welcome black box hypotheses, though not hail them as the culmination of the research process, and we must resist the black box philosophy that shuns mechanismic hypotheses (Bunge, 1964).

In the history of science it has been common for phenomenological and mechanismic hypotheses to conflict over a domain of facts. More often than not mechanismic hypotheses have either replaced or supplemented their phenomenological rivals. Recall the following cases: (*a*) the hypotheses about electromagnetic fields eventually displaced those of action at a distance (or direct inter-particle interaction); (*b*) chemical kinetics, which centers around equations for the rate at which chemical reactions proceed, are supplemented by hypotheses concerning reaction mechanisms (intermediate reactions), which are in turn explained by quantum chemistry in terms of forces and collisions; (*c*) the hypothesis that biopopulations evolve has been supplemented by the hypothesis that evolution proceeds by spontaneous genetic changes, natural selection, geographical isolation, etc.; (*d*) the descriptive generalizations of mentalistic and behavioristic psychology are supplemented or replaced by neuropsychological hypotheses; (*e*) the matrices of social mobility theory are supplemented by hypotheses about the economic, cultural and political "forces" causing job and status drifts.

Needless to say, a mechanismic or dynamical hypothesis is supposed to describe with precision some mechanism or other, not just state that there is, or must be, a mechanism. The latter would be a *programmatic* hypothesis inspiring the search for precise mechanismic hypotheses. Thus Maxwell's equations for the void describe in detail the mechanism of propagation of an electromagnetic field. On the other hand, when Chomsky (1980) and his school claim that the human mind contains innate mechanisms for the generation of grammatical sentences, they state only a programmatic hypothesis, for they do not even sketch such alleged mechanisms. And, since the conjecture of the inborn generative grammar is imprecise, almost any data can be construed as confirming it. Moreover, in a psychobiological perspective the mind cannot contain any mechanisms, for the mind is only a collection of neural processes (Vol. 4, Ch. 4). There are on the other hand neurophysiological mechanisms, in particular mechanisms of formation and dissolution of plastic neural systems. In other words, the psychological mechanisms are neurophysiological. No matter, no mechanism.

What holds for psychological mechanisms holds also for social ones: they are all mechanisms of concrete things—in this case living individuals and functional social systems. For example, the hypothesis that a high rate of interest is a cause of inflation, not a deterrent to it, is true but superficial, for it does not indicate any mechanism. A more complete story, and therefore a better guide to action, is contained in the hypothesis that a high rate of interest decreases the rate of borrowing, which in turn diminishes spending, which decreases production and trade, which decreases employment, which in turn decreases spending, and so on. Incidentally, the central hypothesis of monetarism, namely that the essence of an economy is the flow of money, and therefore controlling money is the most effective way of controlling the economy, is a typically superficial (and false) hypothesis: producers and consumers, together with resources and commodities, not the means of exchange, compose an economy. (More on this in Vol. 6.)

We have declared our preference for mechanismic or dynamical hypotheses over phenomenological or kinematical ones. This preference for depth, though justified, is far from universal. In fact most people prefer black box hypotheses because they are easier to frame, fit to the data, and test. In addition, black box hypotheses have been recommended on philosophical grounds. One of them is conventionalism: If all our assumptions are mere conventions and none represents reality, why not choose the simplest, in particular the kinematical ones? Another is empiricism: If all we can get to know are appearances, why conjecture hidden mechanisms instead of remaining content with hypotheses relating the observable inputs and outputs?

Conventionalism as well as empiricism include then simplicism, or the methodological rule that recommends preferring the simplest of all hypotheses compatible with the data. There are two main variants of simplicism: moderate and radical. The former is Ockham's razor, or the principle according to which "Plurality [of entities, properties, causes, etc.] is not be be assumed without necessity". This variant of simplicism is moderate, for it discourages the invention of fictions but not that of whatever may be needed to explain things. The second variant of simplicism is summarized in the principle of economy of thought, proposed by Avenarius and Mach. According to this principle our aim as cognitive subjects should be to account for our own experience with the least possible intellectual effort. The upholders of this principle used it rather successfully, at the turn of the century, to ignore field physics, attack the nascent atomistics, and try to reconstruct mechanics as a kinematical theory, i.e. dispensing with the concepts of mass and force. We have learned since that simplicity is the seal of ignorance not of wisdom, and simplicism a paralyzing credo: reality is complex and accordingly our representation of it is bound to be complex. (See Planck (1910) and Bunge (1963) for criticisms of simplicism.) Still, the myth of simplicity survives among some philosophers who, even to this day, repeat Mach's dogma that a scientific theory "must provide an economical statement of the regularities which hold between observable phenomena".

3. CHANCE AND CAUSE

3.1. *Probabilistic Hypotheses*

The great majority of hypotheses state associations between putatively objective properties of concrete things, such as atoms, fields, cells, persons, or social systems. There are several kinds of association and correspondingly of association hypotheses. We shall mention the following:

(1) *Qualitative association*, such as "All *A*'s are *B*'s", "Every *A* is accompanied (or preceded or followed) by some *B*", and "The occurrence of *A* prevents that of *B*".

(2) *Statistical association*, such as "*f* percent of *A*'s are *B*'s", "There is a strong positive correlation between schooling and earning", and "There is a strong negative correlation between social participation and crime".

(3) *Functional relation*, as in "Thing *a* reaches (or reached or will reach) place *b* at time *c*", and "The total consumption in any given year is a linear function of the national income in the preceding year".

(4) *Probabilistic association*, as in "The probability that *a* be at place *b* at time *c* equals *p*", and "The probability of an *A* turning into a *B* within the next time period is *p*".

(5) *Descent relation*, such as "All molecules self-assemble from their components", and "Species *A* and *B* have a common ancestor *C*".

(6) *Causal relation*, such as "Gravitation produces acceleration", and "Discontent causes rebellion".

Whereas the hypotheses of the first three types tell us (conjecturally) what goes with what, those of the last two types tell us what comes from what. The probabilistic hypotheses (4), which are a particular kind of functional relations, may represent either regular associations or transformations. The phenomenological (kinematical, black box) hypotheses are all of the types (1), (2), (3), or (4), whereas the mechanismic (dynamical) hypotheses are of the types (4), (5), or (6).

In the research process the various kinds of hypothesis remain distinct but not separate. Thus some correlations suggest or confirm descent or causal hypotheses, and the latter may result from suitably interpreting certain functional hypotheses. This statement calls for some elaboration. Take a random sample from a population, and measure two properties possessed by their members. If the two properties happen to be strongly correlated, it is legitimate (though perhaps false) to conjecture that a suitable partition of the original sample may yield a set of subsamples in each of which the correlation coefficient is maximal, i.e. equal to 1. See Figure 8.4. In other words, a correlation may hide a whole family of functional relations, and be thus just a sign of heterogeneity in the population. (Caution: the converse is not true, i.e. two variables may be functionally related without being linearly correlated. Feller (1968, I, p. 236) offers the following example. Let U and V be variables with the same distribution, hence the same mean, and let $X = U + V$ and $Y = U - V$. Then $\overline{XY} = 0$ and $\bar{Y} = 0$, whence $\rho(X, Y) = 0$. This shows that correlations do not yield general measures of dependence and that they do not generalize functions.)

In the philosophical literature (e.g. Nagel, 1961; Hempel, 1965) statistical generalizations, such as "f percent of A's are B's", and "A and B are strongly correlated", are usually mistaken for probabilistic hypotheses. One reason for this confusion is the empiricist thesis that a probability value is nothing but a long run frequency. This thesis was shown long ago to be

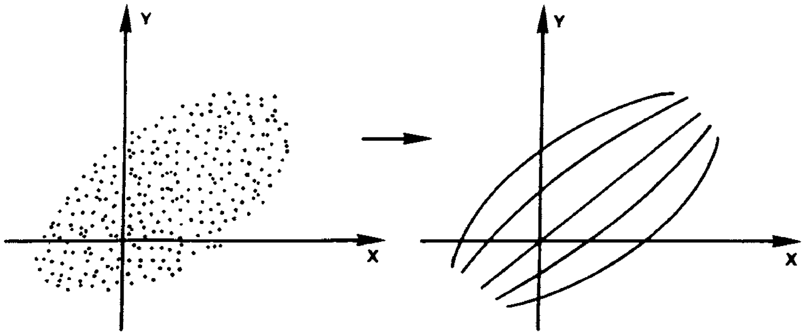


Fig. 8.4. From correlation to functional relation(s).

mathematically false (Ville, 1939). It is also philosophically untenable. For one thing, probability is a theoretical concept whereas frequency is an empirical one. For another, whereas probabilities concern potential states or events, frequencies are of actual states or events. What is true is that some (by no means all) probability statements are tested by confrontation with frequency statements. (See Bunge, 1973a, 1981c.)

Probabilistic hypotheses are theoretical and they involve, either fundamentally or derivatively, probability distributions. A good example of such hypothesis is the Maxwell–Boltzmann distribution of the velocities of the molecules in a gas: Figure 8.5. The operationalist reading of this curve is in terms of the *frequency* with which a molecule will be *found* to have a given velocity. The realist interpretation involves the concept of probability, not frequency, and says nothing about our finding anything: it says that the probability or propensity that a molecule will have a certain velocity is such and such—whether or not the velocity is measured. Considerations of measurement appear only at the test stage, when probability distributions are contrasted with frequency distributions or histograms, such as that of Figure 8.6.

A theoretical probability distribution is a very potent hypothesis, for—when conjoined with the suitable definitions—it yields any number of parameters, such as averages, moments and standard deviations (or scatters). In most applications these parameters are global or collective properties of the population concerned, be it a physical, biological, or social system. Thus the mean velocity of a gas and the average age of an animal population are global (or nondistributive) properties of the whole. However, in the quantum theory every individual entity is assigned a whole

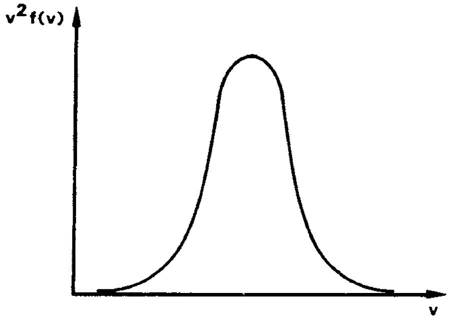


Fig. 8.5. The Maxwell–Boltzmann distribution for the probability that a molecule in a gas in equilibrium has a given velocity.

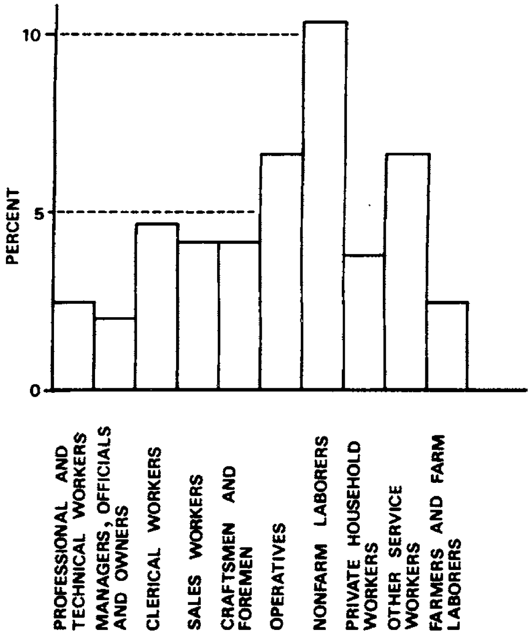


Fig. 8.6. Unemployment rates by occupation in the U.S.A. during the 70's.

set of probability distributions (of position, linear momentum, etc.). In either case some of the parameters calculated from a distribution can be checked by comparison with the data contained in a frequency distribution (histogram).

Whether classical or quantum-theoretical, probabilities are objective (though perhaps falsely posited), not subjective. (Objective probabilities are often called 'propensities', and subjective probabilities 'credences', 'degrees of rational belief', or 'personal probabilities'.) That this is so can be seen from the following considerations. First, all the probabilities occurring in science and technology concern concrete things rather than our beliefs about them. Second, in probability matters our beliefs are expressed by our favoring certain probability distributions over others—hence they are formulated as metastatements such as "I favor such and such distribution". Third, we are supposed to check such hypotheses by contrasting our probabilistic hypotheses, and their consequences, with empirical data. (For details see Bunge, 1981c.)

A probability distribution may be proposed to explain an empirical frequency distribution or it may suggest collecting some statistic. Everyone should feel free to propose probability distributions, whether on the strength of data, or of theoretical considerations, or by sheer guessing. The only conditions are that (*a*) the hypothesis be checked, and rejected if found false, and (*b*) the thing or process in question be, or be assumed to be, random in some respect. The first condition is obvious to anyone save an extreme subjectivist. The second condition is ignored by most of the philosophers interested in probability, as shown by the popularity of the procedure of assigning probabilities to propositions such as hypotheses and data. However, no randomness, no chance for probability.

If proposed prior to the collecting of empirical data, a probability distribution is said to be *prior*. Thus Maxwell and Boltzmann proposed the Maxwell–Bolztmann distribution several decades before it was possible to check it experimentally—and even before the hypothesis of the existence of atoms and molecules was generally accepted. A prior distribution is not something to be accepted or rejected on subjective grounds, but a hypothesis to be put to the test. Guessing prior probability distributions is methodologically akin to guessing any other functional hypotheses; and guessing probability values is no different from guessing distances or ages. All such judgments, whether of probabilities or of sizes, are guided by incomplete data and by hunches (sometimes called 'heuristic rules') and are, of course, subject to error (Tversky and Kahnneman, 1974). There is nothing wrong with such errors in subjective estimates of probabilities, or of anything else, as long as we have the chance of revising them—which is usually the case in science and technology but not so in everyday affairs, where bad mistakes of the sort are often paid dearly.

What is definitely wrong is the subjectivistic (or personalist) doctrine of probability (de Finetti, 1972; Savage, 1972). According to this doctrine probabilities (*a*) express nothing but credences, and consequently (*b*) they are always estimated subjectively. If a scientist or a technologist were to proceed in this way he would be charged with violating the ideal of objectivity. The theoretical physicist, chemist, or geneticist who calculates the probability of the transition of a system from one state to another does so not on a valid hunch but on the basis of law statements such as the Schrödinger equation. And he takes it for granted that, if the scientific theory is correct, the calculated probability quantitates the tendency or propensity of the system to make such a transition.

The science of chance was born only a few centuries ago, and it is not yet taught in elementary school. Consequently we are still unfamiliar with the ideas of randomness, correlation, and their kin. This gives rise to the gambler's fallacy, which may lead an individual to his ruin. There is an even more basic mistake, which is betting in games of chance. Consider the so called "lottery paradox" (Kyburg, 1961), which has elicited the flow of so much philosophical ink. A person ignorant of the nature of chance bets that the (presumably fair) die he is rolling will not come up face *n*, where *n* ranges between 1 and 6. If each of these six statements is taken as true, their conjunction must be accepted as well—but of course it is false, because the 6 possible events are mutually exclusive. The paradox is dissolved on realizing that each of the above statements of the form "The die will not come up *n* when tossed" is groundless. A probabilistic analysis of the game of dice allows one to make only probabilistic statements, such as "In the long run this die will not come up *n* about five out of six tosses". The moral is obvious: Don't bet if you can avoid it.

Another example is the classical finding that most lay subjects are biased towards conjecturing associations (i.e. joint occurrences of properties or events) and neglect negative instances (Smedslund, 1963). Experienced nurses were presented with a 2 × 2 contingency (or presence–absence) table such as this one:

		Disease	
		Present	**Absent**
Symptom	**Present**	*a* 37	*b* 33
	Absent	*c* 17	*d* 13

85% of the nurses estimated that the symptom–disease correlation was strong, for they focused on the quadrant *a* with total neglect of the other three cells. The actual correlation coefficient is negligible, namely $37/70 - 17/30 = 0.038$. Ward and Jenkins (1965) found similar results working with college students presented with contingency tables concerning the association of cloud seeding with rainfall.

3.2. *Causal Hypotheses*

Whereas the statistician is happy if he finds a strong association between two variables, the scientist is unhappy if he cannot explain it, and the technologist if—for want of knowledge or of means—he cannot alter it. For example, sociologists would like to know why schooling and earnings used to be strongly correlated. Was it that increase in schooling produces a higher earning potential or, conversely, that a higher income makes longer schooling years possible?

Statistical correlations can be explained by hypothesizing (*a*) some underlying probability distribution(s), or (*b*) a third factor influencing, or even generating, the observed associated traits, or (*c*) a causal relation between the two variables—or rather their variations. Let us deal with the latter, which are the most popular and at the same time the most reviled by the empiricist tradition, which regards the notions of cause and effect as remnants of metaphysics. (See Bunge (1959) for details on causation and its philosophy.)

The traditional causal hypotheses are of the form *A is necessary for B*. (Sufficiency may require further concurrent causes.) Since the birth of probabilistic theories in science and technology we must reckon also with semicausal probabilistic hypotheses of the basic form *The probability of A causing B equals p*. In these formulas '*A*' and '*B*' stand for events or kinds of events. Things and their properties do not qualify as causes and effects: only changes in properties of concrete things do. (Hence an initial state cannot be said to cause a final state—the more so since the process in question may be uncaused, i.e. spontaneous.) There would be no causation in an unchanging world, for causation is a mode of becoming, not of being.

In general, a functional relation cannot be read as a causal hypothesis without further ado. For example, although the weight of a body is a function of its volume and specific gravity, these cannot be said to cause the former: weight is an effect of the acceleration caused on bodies by a gravitational field. Some functional relations can be interpreted causally, others cannot. Take for instance a functional relation of the form "$y=f(x)$", where x and y are real variables. If x and y happen to represent events, i.e. changes of some kind, then perhaps x can be interpreted as the cause of y or conversely. If not, we may try the derived formula "$\mathrm{d}y=f'(x)\,\mathrm{d}x$", where $\mathrm{d}y$ is the variation in y corresponding to the change $\mathrm{d}x$ in x. A priori it may be that $\mathrm{d}x$ is the cause of $\mathrm{d}y$, or conversely, or that neither causes the other—i.e. that $\mathrm{d}x$ and $\mathrm{d}y$ are merely conjoined or associated. Which of the three interpretations is the correct one must be found out by additional research, either by looking for underlying mechanisms or by experimentally controlling or inhibiting changes in either variable and measuring values of the other.

Take the well-worn case of the ideal gas at constant temperature in a container equipped with a movable piston, and "obeying" the law

"$pV = \text{const.}$". If the piston compresses the gas, the decrease $-\mathrm{d}V$ in volume is the cause of the increase $\mathrm{d}p$ in internal pressure. But if the gas is allowed to push the piston, the decrease $-\mathrm{d}p$ in internal pressure will cause the increase $\mathrm{d}V$ in volume. Both causal relations are covered by the single formula "$\mathrm{d}p.V + p.\mathrm{d}V = 0$" derived from the gas law. In short, functional relations do not become causal hypotheses unless supplemented by causal interpretations. Moreover in some cases such interpretations cannot be appended. For example, Schrödinger's equation in time does not describe the causal evolution of the state function, because the latter is a probability amplitude. And even in elementary mechanics there are problems that have any desired number of solutions (Truesdell, 1974), so that classical mechanics can hardly be said to be a paragon of causal theory.

Moreover certain nonprobabilistic ("deterministic") models involving nonlinear differential equations have nonperiodic solutions that look random (and are miscalled 'chaotic'). That is, the solutions look totally irregular even to the trained statistician, although they do not contain the probability concept and, moreover, they derive from equations representing nonstochastic processes. See Figure 8.7. This is particularly the case with equations with time delays, capable of representing after-effects such as gene expression or the effect of a temporary hump in the birth curve on the market two decades later. This mock randomness is the more puzzling as only equations with time delays can adequately represent causal relations, all of which are supposed to satisfy the principle of antecedence, according to which the effect emerges after the cause. The investigation of

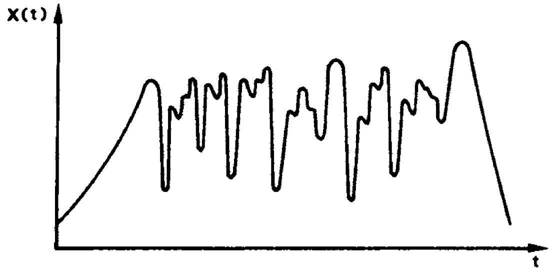


Fig. 8.7. The "deterministic" equation with time delay

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \frac{ax(t-\tau)}{b^n + x^n(t-\tau)} - cx$$

has nonperiodic or "chaotic" solutions for large values of τ and n. See Mackey and Glass (1977). For further examples see R. H. G. Hellman (Ed.) (1980).

this problem has only begun (see, e.g. Hellman, 1980) and it is still to be exploited philosophically.

The above example shows vividly that mere empirical data cannot suggest unequivocally whether causation or mere functional relation, genuine or spurious correlation, or even chance, is at stake. A simpler example is this. Suppose you are given the sequences 0000000000 and 0100111011. You may be tempted to bet that, whereas the first was generated by some causal mechanism, the second was produced by a randomizer. Yet both sequences have the same probability, namely 2^{-10}. In sum, data are not enough to tell the difference between randomness and causation. Nor does intuition, even when accompanied by experience, suffice. We also need some hypothesis about the underlying mechanism—a testable conjecture of course—the more so since certain nonprobabilistic mechanisms give rise to behavior that looks random.

Whereas statisticians are usually content to find correlations, scientists and technologists are interested in discovering whether there is a functional relation, perhaps even a causal one, behind a regression line. This is a tricky problem that a mere contemplation of correlations cannot solve. In fact, although weak correlation disproves conclusively functional relation and causality, strong correlation may suggest either without proving them. Thus the strong correlation between per capita GNP and automobile production can be explained, *prima facie*, in either of the following alternative ways: (*a*) rising GNP causes rises in automobile sales—but then why does it not cause a similar increase in the sales of philosophy books?; (*b*) GNP and automobile sales are two different indicators of industrialization—but then why are they not similarly associated in the industrialized socialist countries?; (*c*) the growth of cities and the inadequacy of mass transportation systems cause the need for more automobiles, a need that can be satisfied by a rising per capita GNP—but then why do the Western European nations exhibit a similar correlation while having generally adequate collective transportation systems? In sum, although strong correlation may suggest causation it does not indicate its direction and, of course, it does not prove it. Monetarist economists have still to learn this lesson, for they like to buttress their hypothesis of the causal efficacy of money with some favorable correlations while disregarding the unfavorable ones. (See Modigliani (1977) for the negative correlation between monetarist policies and prosperity in the post-war U.S.A.)

But why bother with causality if quantum theory, the most fundamental physical theory, is probabilistic? Does this not prove that causality is a myth? This was indeed the impression initially given by the quantum-theoretical revolution, particularly when allied with the positivistic criticism of causality. But on closer examination quantum physics is seen to have causal as well as probabilistic ingredients. Indeed probabilities evolve under the influence of forces—and forces are causes. (More precisely, the form of the probability amplitude depends critically upon the form of the hamiltonian, which represents the forces among the components of the system as well as the forces exerted by the environment on the system.) For example, when a physicist calculates or measures the probability that an atom will scatter an incoming electron within a given solid angle, he determines the propensity of a given event (the two entities coming close together) *causing* another event (the scattering). In short causality has blended with probability.

What holds for physics holds a fortiori for all the other factual sciences. In all of them we find causal hypotheses, sometimes including a chance ingredient, as in the schema "The probability of A causing B is p". Causality, far from having disappeared, is alive and well though now as a component of a richer determinism (Bunge, 1959a, 1982a).

4. REQUIREMENTS

4.1. *Strength and Precision*

We make conjectures all the time but, unless we are fools, we do not assign them all the same status. Upon examination some hypotheses prove to be weaker, or wilder, or less deep, or less testable, or less true, or less useful than others. We should accept only those that pass certain tests—and even so we may have to demote or even discard some of them as new knowledge becomes available. Now, there are several kinds of hypothesis and therefore, presumably, different sets of requirements. Therefore we should start by distinguishing hypotheses with regard to their methodological status.

Our very first distinction is between substantive and nonsubstantive hypotheses. A *substantive* hypothesis is a hypothesis that describes, explains, or forecasts facts of some kind. The phenomenological and mechanismic hypotheses, as well as the probabilistic and causal ones, are substantive. On the other hand the *nonsubstantive* hypotheses are valuational or prescriptive: they guide research and action. We distinguish two kinds of them: methodological and axiological. A *methodological* (or *instrumental*) hypothesis is a tentative statement concerning the methodological status of a statement, a procedure, or a plan—e.g. that a certain formula is, or is not, a suitable axiom or a convenient definition, or that a certain procedure is, or is not, adequate for attaining some goal, or that a certain political tactic is likely to bring victory. And an *axiological* (or *valuational*) hypothesis is a value judgment—e.g. that a given theory is all round better than another, or that a certain action or a given life style is good. Obviously, we make hypotheses of all these kinds in all walks of life, from pure mathematics to mass murder.

We should impose two requirements on hypotheses of all kinds, whether substantive or nonsubstantive. Of two rival hypotheses we should prefer the better grounded or justified, and the one belonging to the wider conceptual system. (This is a norm, not a hypothesis.) The former condition because rationality entails, among other things, the rejection of authority, even that enjoyed by success. The second condition because the wider the conceptual system the more propositions can check the given hypothesis. More on these conditions later on.

Aside from those two conditions, we treat differently hypotheses of different kinds. We prefer the strongest, more precise, deepest, more systematic, and better confirmed substantive hypotheses. These criteria are epistemic, not utilitarian. On the other hand what we expect from nonsubstantive hypotheses is effective guidance. Thus from methodological hypotheses we expect that they help us solve methodological problems of some kind, such as the organization or test of a body of propositions, the gathering of data, the design of artifacts, the outlining of plans, or what have you. And from axiological hypotheses we expect that they help us choose or do the best or avoid the worst. In the remainder of this section we shall deal only with substantive hypotheses, and shall do so in several respects.

From a logical point of view hypotheses may be ordered in respect of strength. A formula A is said to be *logically stronger* than a formula B if, and only if, A entails B. (In turn A entails B, or $A \vdash B$, only when the conditional "If A then B" is logically true, or holds for all interpretations of A and B.) Thus p is logically stronger than "p or q", where q is an arbitrary proposition, i.e. one not necessarily related to p; "p & q" is stronger than either p or q; and an "all" statement is stronger than the corresponding singular statement, which in turn is stronger than the corresponding "some" statement.

Paradoxically, the strongest of all formulas are the logically false ones, i.e. the contradictions, for they entail anything. On the other hand the tautologies, or logically true formulas, are the weakest of all, for they entail nothing but further tautologies, and they follow from no premises. Logic, in short, is self-validating and self-contained—in fact it is the only such epistemic field. For this reason it cannot be invalidated by any factual discoveries. However, this does not entail that logic is detached from the rest of human knowledge. Far from it, logic underlies (is presupposed by) all rational discourse. In particular, every consistent theory contains logic as a proper part. More on logic in Vol. 6.

Whereas in logic we seek to establish the weakest formulas, in all other fields of research we steer a middle course between tautology and contradiction. However, with regard to logical strength there is a marked difference between the course taken in the theoretical branches of a science and that taken in its empirical (laboratory or field) ones. (Except for the formal sciences, all the others have both theoretical and empirical branches. Thus physics is composed of theoretical and experimental physics, and biology of mathematical biology and experimental biology.)

Indeed, in theoretical science we should prefer the *strongest* noncontradictory hypotheses compatible with certain general principles as well as with the relevant examples (in the case of mathematics) or empirical data (in the case of factual science). On the other hand, laboratory and field research should seek to establish the *weakest* nontautological formulas compatible with the most general principles of science, such as the principle of conservation of energy. Such are the data contained in laboratory protocols and in the notebooks of field researchers. They are in principle corrigible in the light of further data or of well-confirmed hypotheses but they do not replace or entail hypotheses.

Let us consider *precision* next. Like concepts, hypotheses come in various degrees of precision, from the ones couched in ordinary language to those couched in a mathematical language. Thus compare "There must be some thing, somewhere, that is doing something to this thing", with "Set A is included in set B". Of all the ordinary language expressions, counterfactuals break the record of imprecision. Thus the sentence 'If A were the case then B would happen' can be interpreted in a number of ways, such as "If A then B, but A is false", and as an injunction to bring about A so as to attain B. This is why counterfactuals do not occur in scientific theories and why it is wrong-headed to try and build exact theories about them.

The most precise formulas are of course those containing exact concepts. But even among mathematical formulas we can distinguish different degrees of precision. Thus an equation is more precise than the corresponding inequality, and a quantitative formula more precise than the corresponding qualitative ones. Compare "f decreases monotonically" with "For all real $t \neq 0, f(t) = 1/t$". There are infinitely many functions that fit the former description, only one that fits the latter. Consequently the chances of being right are greater when proposing an imprecise hypothesis than when proposing a precise one. But scientists are not particularly interested in cheap truths, just as moralists are not interested in passing through wide doors. Other things being equal they prefer the more precise hypotheses because they "say" more.

The preference for precision over imprecision is justifiable only if there is an option. When there is no option we must settle for imprecision—for the time being. This is unavoidable, nay desirable, at the beginning of a research project, when we formulate *working* (or *programmatic* or *schematic*) hypotheses that guide the investigation. Thus the hypothesis that every mental phenomenon is (identical with) the specific activity of some plastic neural system, summarizes an entire research program in physiological psychology—a very fertile one indeed.

Three kinds of working or programmatic hypotheses are of particular interest: functional schemata, existential conjectures, and negative ones. A *functional schema*, such as "Y is some function of X", cries for precision and suggests the search for the exact functional relation between X and Y. It is thus part of a research project aiming at substituting a precise function (e.g. the exponential one) for the indeterminate function symbol F in "$Y = F(X)$". Or take the *existential* hypothesis "There is one genotype for every phenotype". This (partially true) working hypothesis inspires the whole of genetics: in fact, a central task of geneticists is to map gene complexes into phenotypes—a task that is unlikely to be ever completed, so many are the biospecies and their traits. Finally, consider the *negative* hypotheses that there are no microphysical inertial systems, and that the speed of light *in vacuo* fluctuates around c instead of being constant. These suspicions could generate a research line resulting in a generalization of relativistic physics. (Being negative hypotheses they can belong to no theory although they may suggest building a theory complying with them.) In sum, we cherish working hypotheses as parts of research projects. The advancement of knowledge consists, among other things, in a march from programmatic hypotheses to precise substantive ones and, in general, a progression from imprecision to precision.

4.2. *Systemicity and Scrutability*

A third property of hypotheses we must consider is their systemic status, i.e. whether or not they belong to an organized body of knowledge, such as a classification or a theory. Most of the hypotheses we formulate in daily life are stray. Thus we may conjecture that a colleague's dark glances at us are due to jealousy, although we have no theory of jealousy and no reliable jealousy indicators or objectifiers. In advanced science and technology we prefer *systemic* hypotheses, i.e. conjectures belonging to conceptual systems, and distrust the nonsystemic or *stray* ones. The advantages of systemic over stray hypotheses are obvious: (*a*) the hypotheses in a system can join to generate (logically) further hypotheses, and (*b*) systemic hypotheses are supported not only by whatever empirical data may be compatible with them, but also by the other components of the system. See Figure 8.8.

Stray hypotheses may concern (refer to) a large or a small domain of facts. The stray hypotheses that account only for a small subset of the total set of facts it addresses are called *ad hoc* (for that). Here are three examples of stray hypotheses that are also *ad hoc.* (*a*) "Criminals are the product of society". (The same society "produces" a majority of noncriminals.) (*b*) "Underground and cave dwellers have lost their sight for lack of use of their eyes". (Not all such animals are blind. Besides, genes are presumably not affected by luminosity.) (*c*) "The dinosaurs became extinct because their brains were too small to cope with a complex environment". (The dinosaurs had been remarkably successful, i.e. prolific, before the end of the Cretaceous period—at which time many other species became extinct as well.)

Whereas some *ad hoc* hypotheses are proposed to "save the appearances", i.e. to cover or even explain some group of facts, others are

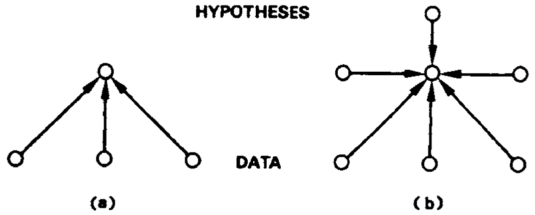


Fig. 8.8. The systemicity of hypotheses. A stray hypothesis (*a*) is supported at best by the data favorable to it. A systemic hypothesis (*b*) is supported also by the other hypotheses in the same conceptual system. The arrows indicate support, not inference.

introduced to save further hypotheses endangered by unfavorable evidence. We call the latter *protective* hypotheses. They are not all alike: whereas some may turn out to be true, others are not even scrutable. We call *bona fide* protective hypotheses those scrutable conjectures that are designed to save other hypotheses, and *mala fide* those which, being introduced for the same reason, are inscrutable. Whereas the former can in principle be shown to be true or false to some extent, the latter cannot.

The hypothesis that the naturally occurring elements are mixtures of different isotopes was initially framed to save the atomic hypothesis: it was testable and eventually found true. Likewise the hypothesis that an organism does not respond in the same way every time it is subjected to a given stimulus, for it keeps changing its internal state, was originally introduced to save stimulus–response psychology but it turned out to be true. Both are examples of *bona fide ad hoc* protective hypotheses. On the other hand the psychic's claims that his failures are due to somebody's hostility inhibiting his paranormal abilities, and the psychoanalytic fantasy that most people do not exhibit their Oedipus complex because they have repressed it, are paragons of *mala fide* protective hypotheses.

Mala fide protective hypotheses are peculiar to everyday reasoning and pseudo-science. On the other hand *bona fide* protective hypotheses are very common in science, technology, and the humanities. Sometimes they are introduced to discard data that collide with well established theories. (Imagine an engineer rejecting classical mechanics because his bridge collapsed.) At other times *bona fide* protective hypotheses are framed to shield new theories from rash criticism. (Imagine Darwin giving up his theory of evolution because of the gaps in the fossil record.) There is nothing wrong with a protective hypothesis as long as it is *bona fide*, i.e. independently testable. (For further examples see Bunge, 1967a, Ch. 5, Section 5.8, and 1973b, Ch. 2, Section 1.) We shall come back to this question in Ch. 11, Section 2.1.

The fourth and last property of hypotheses we need to look into in this chapter is scrutability; truth and utility will be examined later on. A substantive factual hypothesis can be said to be *scrutable* or *testable* if, and only if, (*a*) it refers exclusively (whether truly or falsely) to real (material) entities, and (*b*) it can be checked for truth by contrasting it with data or with the bulk of accepted knowledge. Otherwise the hypothesis is inscrutable or untestable. A hypothesis that is both inscrutable and stray (nonsystemic) may be called *wild*. Typically, pseudo-scientific speculation is wild.

Scrutability, a methodological predicate, does not coincide with intelligibility, a psychological one. A bold new hypothesis may be both scrutable and hardly intelligible to but a few initiates. Nearly every profound innovation is baffling: "When the great innovation appears, it will almost certainly be in a muddled, incomplete and confusing form. To the discoverer himself it will be only half-understood; to everybody else it will be a mystery. For any speculation which does not at first glance look crazy, there is no hope" (Dyson, 1958, p. 80). There is no great and interesting research, in any field, without speculation; the only condition we must impose on it is that it be scrutable, not wild.

Some sciences are and will probably remain largely speculative. Cosmology, palaeontology and history are among them. Take for instance the speculations of palaeontologists. Far from being wild they are guided by (*a*) fossils, (*b*) associated remains (e.g. those of plants and prey), (*c*) considerations of homology with living organisms, (*d*) geological data and hypotheses, and (*e*) general biological and even purely mechanical principles concerning, e.g., posture and locomotion. For example, it is now possible to argue for the hypothesis that all of the great dinosaurs were water dwellers, as well as for and against the fantastic (but not wild) conjecture that dinosaurs, unlike their modern successors, were warm-blooded and therefore had a quick metabolism. Thus it is now possible to cast "a cold look at warm-blooded dinosaurs"—the title of a recent scientific symposium (Thomas and Olson, 1980).

To conclude. When considering hypotheses for test or adoption we should generally prefer depth to superficiality, logical strength to weakness, precision to inexactness, systemicity to isolation, and scrutability to untestability. However, imprecise conjectures are acceptable if they are promising research guides, i.e. programmatic hypotheses. And superficial and relatively weak hypotheses may be unavoidable if we need quick practical results. What we cannot accept under any circumstance, in any walk of life, are inscrutable hypotheses, particularly of the protective kind, for they are the mark of dogmatism.

5. Concluding remarks

Dogmatic rationalists, intuitionists and radical empiricists have no use for hypotheses: the former own indubitable a priori axioms, the intuitionists are endowed with a special direct insight (some of them even with what Husserl called *Wesensschau*, i.e. vision of essences), and the radical empiricists, more modest, are content with perceptions and images. Even the great Russell, during his logical atomism period, proposed to eliminate what he called 'inferred entities' (i.e. hypothesized objects) in favor of constructions out of certified, in particular perceptible, entities. He proposed the rule "Whenever possible substitute constructions out of known entities for inferences to unknown entities" (Russell, 1914). And he applied this maxim to the analysis of matter, which he dissolved into phenomena. In a similar vein, operationists and members of the Vienna Circle attempted to characterize electrons and other transempirical entities, not in terms of their objective properties—the way the theories of matter do—but in terms of pointer readings or some other empirical operations. Fortunately scientists did not follow up this anthropocentric program—except of course during their occasional forays into philosophy. They went on framing and checking increasingly sophisticated hypotheses about hypothesized (but supposedly real) entities such as fields and subatomic particles, extinct organisms and societies, the interior of stars and of brains, and so on. Humanists and technologists behave similarly: far from avoiding conjectures they frame them all the time; they try to avoid only wild hypotheses.

Hypothesizing is just as necessary as gathering data, because to investigate is to handle problems, and to solve a problem is to frame a hypothesis, or find a set of data, that may satisfy the condition (generator) of the problem. Hypotheses fulfill a number of functions in ordinary life, science, technology, and the humanities. Some of them summarize experience and others generalize it. Some allow us to deal with real or presumptively real but imperceptible facts, still others trigger reasonings or fill gaps in them. Finally, others guide (or block) entire research plans, in particular the design of observations. The services hypotheses render are so many and important that any attempt to dispense with them would be as silly and ineffectual as trying to do without any data.

That the ultimate components of the universe are fields and particles, and all of these are restless; that the sun and the planets are roughly spherical and compose a system; that cells are the units of life and all modern organisms descend from others that lived long ago; that every society is composed of an economy, a polity and a culture—however primitive—and that these subsystems interact with one another vigorously—all these are hypotheses, not data. In fact most of our scientific knowledge is hypothetical, though of course not all conjectures are equally tentative: some of them—such as the ones we have just listed—are as well established as the firmest of data. They are supported not only by data but also by further hypotheses belonging to some conceptual system or other. Conceptual systems is what we shall study next.

CHAPTER 9

SYSTEMATIZING

All humans crave for knowledge, and not just any old heap of bits of information but systems, i.e. sets of epistemic items possessing unity and coherence. The easiest way to obtain both unity and coherence is through radical simplification and reduction to a few basic ideas. This is the way nonscientific ideologies are fashioned.

Simplification is unavoidable and reduction is desirable—up to a certain point, beyond which they involve inadmissible distortions. Reality is much too varied and changeable, creativity much too incessant, and rationality much too demanding, to tolerate simplistic systems. We need conceptual systems capable of accounting for diversity and at the same time possessing the desirable properties of unity and consistency. These systems are classifications and theories.

Despite the craving for conceptual systemicity, there is still some distrust for classification and theorization. We often say we do not like being pigeonholed, and that a gram of practice is worth a ton of theory. This distrust has several sources: sheer anti-intellectualism, misunderstanding of the nature of classification and theory, and ignorance of their role in civilized life. In this chapter we shall study classification and theory, without which man would lead a rather brutish existence, for they help us understand the world and transform it.

We distinguish different degrees of epistemic systemicity or cohesion: the mere collection of bits of information, the context, the classification, the doctrine, the theory, and the system of theories. A catalogue, such as a telephone directory, is an instance of an unstructured set of epistemic items. A context is a more or less loosely knit collection of propositions referring to a single domain. Stories, flow diagrams, wiring diagrams and organigrams are instances of contexts. A context need have no logical structure, so that knowledge of one item of it does not allow one to infer anything about other members. But, by virtue of its definite reference, a context has referential or ontological unity. In Vol. 1, Ch. 2, Section 3.4, we characterized a conceptual context as an ordered triple $\langle P, Q, R \rangle$, where P is a set of propositions, Q the set of all predicates occurring in P, and R the set of referents of the predicates in Q.

Classifications, doctrines, theories (in particular models) and systems of theories are contexts possessing logical as well as referential unity. A classification is a system of propositions concerning items of some kind (e.g. organisms) and their inclusion relations (e.g. "Species S is included in genus G"). A doctrine is a context such that some of its propositions are logically related to others: it need not constitute a hypothetico-deductive system and it need not even be consistent. A theory is a hypothetico-deductive system, i.e. a system of propositions every one of which is either a premise or a logical consequence of a group of premises. And a system of theories is a collection of mutually consistent theories and of bridges among them.

1. CLASSIFICATION AND THEORY

1.1. *Classification*

It seems that all sentient organisms are capable of perceiving certain types and patterns—visual, auditory, tactual, gustatory, olfactory, spatial, temporal, and others. Now, a *type* is whatever is shared by a mass of apparently unrelated items, and a *pattern* is whatever is invariant in a mass of items. Thus all human faces, however great their individual differences, have the same type or general shape; and all waves fit the same general pattern even though they differ in amplitude and frequency. Since we can perceive types and patterns, by the same token we can discard individual differences and thus throw some information away. We do this either from the start or as a result of analysis. In the former case we grasp types and patterns in a synthetic way, sometimes intuitively; in the latter case we conjecture, test, and conjecture again until we grasp type or pattern. In this section we shall study typification, i.e. the formation of types and systems of types.

We saw in Ch. 5, Section 1.2 how classes are formed, namely by grouping together individuals that share certain properties, such as having the same mass or the same descent, even if they differ in all other respects. A single attribute A and its complement not-A allow one to make black or white statements such as "c is an A" and "c is a non-A". A pair of attributes, A and B, allow us to form four different propositions: "c is an A and a B", "c is an A and a non-B", "c is a non-A and a B", and "c is a non-A and a non-B"—i.e. the components of a 2×2 contingency table. In general, for n attributes we may construct 2^n propositions for any given individual. Each such set is called a *Boolean partition*. We use such qualitative partitions every day.

For certain purposes we must use quantitative predicates, such as distances and probabilities, which allow us to form infinitely many propositions, and so infinitely fine partitions. Think, e.g. of the infinitely many distances a moving point can be from a fixed center. We call these *Cartesian partitions*. See Figure 9.1. Because a Cartesian partition contains infinitely many cells, it provides a far richer representation of reality than any Boolean partition—so rich, in fact, that it incorporates information that cannot be obtained by observation and measurement. (Any run of measurements yields only a finite number of numerical values.) It is thus a formidable analytic tool and, by the same token, an inadequate categorization tool, for categorizing is lumping as well as splitting. Let us take a closer look at classification, whose principles are often misunderstood by its most conspicuous users, namely the biologists.

A *classification* of a given collection of individuals—be they concepts, concrete things, events, or what have you—is a conceptual operation with the following characteristics:

(i) each member of the original collection is assigned to some *class*;

(ii) each class is composed by some of the *members* of the original collection, and no class is composed of subclasses;

(iii) each class is a *set* the membership of which is determined by a predicate or a conjunction of predicates;

(iv) each class is *definite*, i.e. there are no borderline cases—which is ensured by employing only definite or exact predicates and avoiding vague ones such as "young" and "distant";

(v) any two classes are either mutually *disjoint* (i.e. have no members in common) or one of them is *included* (contained) in the other: if the former

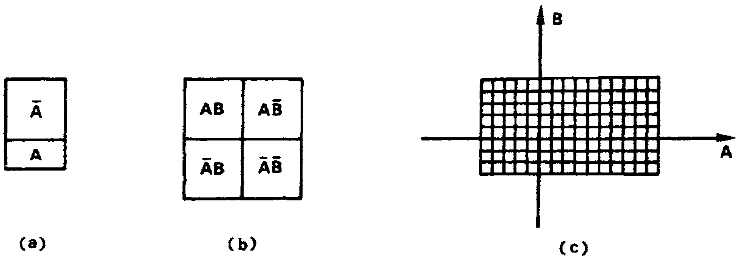


Fig. 9.1. Categorizations. (*a*) Dichotomy. (*b*) A Boolean partition on the basis of two qualitative predicates. (*c*) A Cartesian partition on the basis of two quantitative predicates.

they are said to belong to the same *rank*, otherwise to different ranks;

(vi) only two relations are involved: the *membership* relation ($\in$) holding between the individuals of the original collection and the first rank classes, and the inclusion relation ($\subseteq$) that relates classes of different ranks (e.g. species to genera);

(vii) every class of a rank higher than the first equals the *union* of some or all classes of the immediately preceding rank (e.g. every genus equals the union of its species);

(viii) all the classes of a given rank are *pairwise disjoint* (do not intersect), so that no item in the original collection belongs to more than one class of the same rank;

(ix) every partition of a given rank (e.g. every grouping of the original collection into species) is *exhaustive*, i.e. the union of all the classes in a given rank equals the original collection;

(x) a classification violating any of the above conditions must be repaired or given up.

Clearly dichotomies, such as existent/nonexistent, satisfy all of the above conditions. But they are trivial classifications because they involve a single rank, whereas nontrivial classifications are "vertical" as well as "horizontal". This requires seizing on more than one property. (For example, two attributes, P and Q, elicit the formation of two species, namely $S_1 = \{x|Px\}$ and $S_2 = \{y|Qy\}$, and one genus, namely $G = \{z|Pz \text{ or } Qz\} = S_1 \cup S_2$.) Now, unless the two predicates are mutually incompatible, such as sitting down and running, they may induce the formation of partially overlapping classes. Therefore seizing on arbitrary predicates will not produce a classification proper except by accident.

The only method for producing a classification proper, i.e. a categorization satisfying the nine conditions listed above, is the one based on *equivalence relations* such as those of congruence, homology, common descent, and same age bracket. In fact every such relation splits the set on which it is defined into mutually disjoint (exclusive) and exhaustive (inclusive) sets. These sets are called *equivalence* (or *homogeneous*) classes, and the family of such classes the *partition* of the given set S by the given equivalence relation $\sim$, or $P = S/\sim$. See Figure 9.2a. For example, T shirts for men come in three different sizes, so that the relation "the same size as" ($\sim$) partitions T thus: $\mathrm{T}/\sim\, = \{\mathrm{S},\mathrm{M},\mathrm{L}\}$, where $\mathrm{S} \cap \mathrm{M} = \emptyset$, $\mathrm{S} \cap \mathrm{L} = \emptyset$, $\mathrm{M} \cap \mathrm{L} = \emptyset$ (condition viii above), and $\mathrm{T} = \mathrm{S} \cup \mathrm{M} \cup \mathrm{L}$ (condition ix). In this case the species are S, M and L, and the genus coincides with the original set T. A more sophisticated example is the formation of clades on the basis of homologies such as that between the forelimbs of man, seal, and bat.

Obviously there are as many partitions as equivalence relations. Thus when classing fossil vertebrates one may focus on the conformation of the hip bones, or the possession or lack of a predentary bone in the lower jaw, or the conjectured mode of locomotion, or the conjectured habitat, etc. Each equivalence relation induces its own partition of the given set. Consequently using n different equivalence relations $\sim_1, \sim_2, \ldots \sim_n$, we obtain n partitions $P_i = S/\sim_i$, some of which may be identical. This family of partitions may be pictured as a stack of disks, every one representing the same original pie cut in different ways. See Figure 9.2b.

Once a first partition $P_1 = S/\sim_1$ of the original collection has been effected, we have completed the first rank of the classification: it consists of

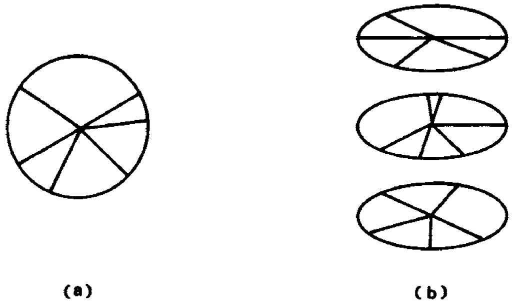


Fig. 9.2. (*a*) Partition of a collection by an equivalence relation. Each sector (subset) is homogeneous in the given respect, and the union of all the sectors equals the original set. (*b*) A family of partitions of a given set.

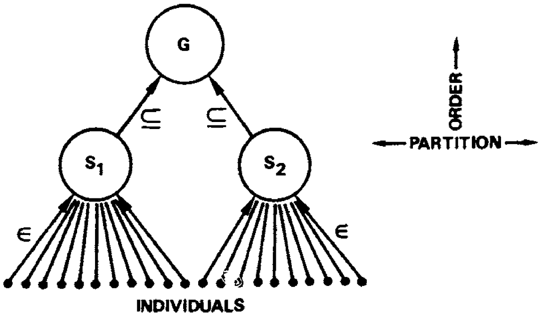


Fig. 9.3. Individuals, species, and genera. Each individual belongs to (∈) some species (lowest order taxon), and each species is included (⊆) in a genus.

species (or subspecies, or varieties). These species may be grouped in turn by using a second equivalence relation $\sim_2$, this one defined on the family P_1, i.e. $P_2 = P_1/\sim_2 = \{G_1, G_2, \ldots G_m\}$. The members of this partition are called *genera*. A third equivalence relation, this one defined on P_2, will induce the formation of the third rank of the classification, namely that composed by *families*—and so on. See Figure 9.3.

To recapitulate, these are the steps in a classification process:

0. A set S of individuals is given or produced.
1. Find a "significant" equivalence relation $\sim_1$ on S.
2. Partition S by $\sim_1$, i.e. form the collection of species $P_1 = S/\sim_1 = \{S_1, S_2, \ldots S_n\}$.
3. Find a "significant" equivalence relation $\sim_2$ on P_1.
4. Partition P_1 by $\sim_2$, i.e. form the family of genera $P_2 = P_1/\sim_2 = \{G_1, G_2, \ldots G_m\}$, such that every member of it be the union of some (one, two or more) species.
5. Repeat the process for the higher ranks as far as necessary or convenient.

One may also invert the process, i.e. start by partitioning the original collection into genera, or even higher rank classes, and proceed to distinguishing subsets within them. For example, the fisherman probably starts by partitioning all fish into edible and inedible—an artificial partition for it is anthropocentric not ichthyocentric. The equivalence relation he uses is "as edible as". He may then proceed to divide the class of edible fish into those which are easy to fish and those which are not. This second equivalence relation, namely "as easy to fish as", is said to *refine* the former. To take a more interesting example, suppose we start with the order primates, and distinguish within it the family of hominids, which in turn includes the genus *Homo*, which in turn includes our own species *Homo sapiens sapiens*. Implicit in this classification are three different equivalence relations that are not always spelled out explicitly. (In fact most biologists partition and order in a rather intuitive way, not caring much for taxonomy, or the methodology of classification. They often pay for this with mistakes or barren controversies. See, e.g. Ghiselin (1981) and the ensuing discussion.) As we descend from the higher to the lower rank classes we do so with the help of finer and finer equivalence relations.

Every classification has two dimensions: the horizontal one attached to the relations of membership ($\in$) and equivalence ($\sim$), and the vertical one linked to the inclusion relation ($\subseteq$). Every good classification is then a trivial example or model of elementary set theory. But not all classifications satisfy every one of the nine conditions we listed a while ago. In particular, some classifications are made with the help of vague or inexact predicates, such as color and shape predicates. And others, actually the great majority, are not exhaustive for want of information. (We keep discovering, inventing or making objects belonging to formerly unknown species.)

Even if a partition is performed with the help of an exact predicate, and therefore an exact equivalence relation, the result may be unsatisfactory because the predicate may be superficial and therefore incapable of bringing about the unity or systematicity we want. For example, the original nomenclature of biocatalyzers (enzymes) was based on the equivalence relation "acting on the same molecule". I.e., if X names the molecule on which enzyme Y acts, then the name of Y is *X-ase*. (Thus arginase catalyzes the hydrolysis of arginine to ornithine and urea.) As the number of known enzymes climbed up to the thousands this nomenclature became impractical. The International Enzyme Commission produced a systematic partition, hence nomenclature, based on a functional description of enzymes. In other words, the new equivalence relation is "having the same function". For example, all the enzymes that catalyze hydrolises are called *hydrolases*, and those that isomerize, *isomerases*.

A clear case of superficiality is that of *artificial* partitions, i.e. those based on traits that, far from being primary and intrinsic, are secondary or concern our relation to the objects concerned (e.g. their external appearance or usefulness). Inevitably, empiricism and pragmatism enjoin us to stick to artificial classifications. And nominalists do not care what classification one may adopt for, according to nominalism, a general sign, such as 'human', does not designate a natural kind but denotes severally the individuals it points to. In other words, according to nominalism the *fundamentum divisionis* of every classification is arbitrary, without a real counterpart: every classification is artificial, and so no classification is better than any other one.

On the other hand, realism recommends natural classifications and, among these, prefers those based on deep or essential properties, in particular laws. (For the notion of a natural kind see Vol. 3, Ch. 3, Section 3.3, and for a defense of nomological essentialism, *ibid.* Ch. 2, Section 4.2.) Essential properties are the basis of essential equivalence relations, which in turn elicit essential or basic partitions—as opposed to accidental or superficial ones. In short, classifications come in several depths, and we should prefer the deepest of all for being the more realistic. (For the notion of depth recall Ch. 8, Section 2.2.) This does not entail that the classes themselves can be more or less real: every class is a concept, but whereas some concepts represent reality others do not. Thus, whereas the set formed by this book and the reader's last dream does not constitute a natural kind, all the bits of copper, and all the hominids and their descendants, do.

Now, mere perception is bound to lead to superficial partitions. Deep partitions call for hypotheses and, in particular, law statements, i.e. formulas about patterns. And, because law statements belong (by definition) to theories, if we want deep classifications we need theories, the deeper the better. Good examples of the power of theory to inspire deep classifications are contemporary biological systematics (based on the theory of evolution), the periodic table of the elements (based on the atomic theory), the classification of hadrons based on the quark model, and the classification of materials based on their constitutive relations or specific laws.

Classing and theorizing are then mutually complementary activities. Categorizing precedes theorizing if only because every theory is about some category of objects. In turn, theory allows one to refine the coarse and shallow pretheoretical classifications. Moreover a classification *is* a theory of a kind. Indeed, unlike the propositions in a catalogue, those in a classification are logically related: they compose a system with logical unity. Thus from the statements that all mammals are vertebrates, and all vertebrates animals, it follows that all mammals are animals. To be sure this piece of knowledge is not a surprising new theorem but was available from the start, but the point is that it is a systemic proposition, not a stray one. Into a classification we pour all we know about certain kinds, and from it we expect no new knowledge. Classifications summarize and order available knowledge, and thus comply with the requisite of empiricism, which prohibits speculation.

Classifications may be dubbed *finite theories*, in contrast to ordinary theories, which contain infinitely many propositions, so that by investigating them we are bound to obtain propositions we did not know before. (See the next section.) For the same reason classifications have a very limited explanatory power. For example, the statement that men and monkeys are primates explains our similitude to our cousins, but only because this similarity was used to construct the order of primates. And what predictive power a classification may have is strictly heuristic, for it stems only from the gaps (e.g. missing links) in it rather than from any explicit statements. For example, the theoretical classification of hadrons based on the quark model predicted the existence of the omega-minus particle, which was eventually found in 1964. A complete classification of known items has no predictive power at all. If we want prediction and explanation we must progress from classification to theory. This we shall do anon.

1.2. *Theory*

From a logical point of view a theory is a logically organized set of statements concerning objects of some kind, i.e. it is a systematic context. At the beginning of this chapter we recalled that a context is defined as an ordered triple of sets: statements-nonlogical predicates occurring in the latter-domain (reference class) of such predicates, or $\mathscr{C} = \langle P, Q, R \rangle$ for short. Since a theory is a logically organized context, it can be characterized as a quadruple $\mathscr{T} = \langle P, Q, R, \vdash \rangle$, where $\vdash$ is the entailment relation that glues the members of P into a system. More precisely, a theory is a *context closed under deduction*, i.e. such that every statement in it is either a premise or a deductive consequence of a set of premises.

If the theory concerns exclusively conceptual objects, such as numbers or philosophical ideas, it is said to be *formal*. If the domain or reference class R of a theory contains factual items, such as molecules or societies, it is said to be *factual*. (Whereas a formal theory contains no reference to factual items, a factual theory may contain not only factual statements but also statements about properties of its own concepts. Thus a theory of change will contain mathematical statements concerning the properties of the functions representing changing properties.) The root difference between a formal and a factual theory lies in the domain R of individuals, or reference class. This difference is semantic and it will occupy us in Section 2.2. In the present section we shall focus on the structure of theories.

That a theory is a logically organized context means that (*a*) a theory is "based on" (presupposes or contains) some theory of logic, usually the predicate calculus with identity, and (*b*) whereas some formulas of the theory are premises, others are logical consequences of some such premises, or even consequences of such consequences—all of them derived with the help of the underlying or presupposed logic. The premises of a theory are either postulates (axioms) or definitions. The consequences of the postulates or definitions are called 'theorems', and the special cases or immediate consequences of axioms, definitions or theorems, are called 'corollaries'. The postulates and their consequences "say" something about the referents of the theory (or members of its domain R), or some

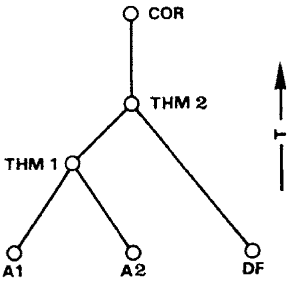


Fig. 9.4. Finite fragment of a theory based on two axioms and one definition. The total theory cannot be pictured because it is a graph with infinitely many nodes (formulas) and edges (logical relations).

concepts of the theory (members of the set Q of predicates), or about relations between concepts and referents. On the other hand a definition links two or more concepts of the theory. See Figure 9.4.

Example 1. Theory of partial order

$A0$ The predicate calculus with identity.
$A1$ R is a nonempty set.
$A2$ For all x in $R, x \leq x$ (reflexivity).
$A3$ For all x and y in $R, x \leq y \,\&\, y \leq x \Rightarrow x = y$ (antisymmetry).
$A4$ For all x, y and z in $R, x \leq y \,\&\, y \leq z \Rightarrow x \leq z$ (transitivity).
Df1 R is *partially ordered* = R satisfies the above axioms.
Df2 R is a *chain* = $\leq$ holds between any two members of R.
Df3 For all x and y in $R, x < y = x \leq y \,\&\, x \neq y$.
Thm1 For all x, y and z in $R, x \leq z \,\&\, z \leq y \,\&\, x = y \Rightarrow x = z$.
Thm2 For all x in $R, \neg(x < x)$.
Cor For all x, y and z in $R, x < y \,\&\, y < z \Rightarrow x < z$.

This is a formal theory, although it can be interpreted in factual terms, e.g. by letting R be the set of bodies and $\leq$ the relation of being less heavy than, or R the family of nations and $\leq$ the relation of being less developed than. It is so simple and general a theory that it underlies (is presupposed by) a host of richer theories, particularly in abstract algebra. By introducing further concepts, such as those of immediate successor, first and last element, greatest lower bound, and least upper bound, richer abstract theories can be built, such as lattice theory and filter theory.

The zeroth axiom $A0$ constitutes the *formal background* of the theory, which in this case is nothing more than ordinary predicate logic with identity. That and the other four axioms and the three definitions compose the *foundations* of the theory. In this case all the axioms concern members of the reference class and their mutual relations. The theory does not contain any semantic axioms (meaning assumptions), i.e. hypotheses about the relation between the specific predicate of the theory, namely $\leq$, and its referents or members of R. In other words, the theory is strictly formal. Its interpretation in factual terms would require the addition of two semantic postulates, one for R, the other for $\leq$—e.g. that R is a set of organisms and $<$ the relation of descent.

Example 2. Theory of exponential growth

*A*0 Ordinary predicate logic with identity plus elementary infinitesimal calculus.
*A*1 R is a nonempty set.
*A*2 F is a real valued function on the cartesian product of R by T, where T is a subset of the real line.
*A*3 F is differentiable with respect to its second argument.
*A*4 The rate F of change of F with respect to its second argument is proportional to F itself, where the proportionality constant is a positive real number depending on the individual (i.e. it is a real valued function on R).
*A*5 'T' represents time.
*A*6 'F' represents the intensity of property P of things of kind R.
Df The relative rate of change of F equals $\dot{F}/F$.
*Thm*1 The relative rate of change of F is constant. (By A4 and *Df*.)
*Thm*2 F grows exponentially. (By *A*4 and certain principles of the infinitesimal calculus.)

The formal background of this theory is far richer than that of the previous theory for, in addition to logic, it presupposes a bulky chapter of mathematics. Thus, because F is assumed to be a continuous function, it has infinitely many values, one for each time and each value of the relative rate of change. That is, *Thm*2 is actually a nondenumerable infinity of formulas of the form "$F(t) = F(0) \exp(kt)$", where $F(0)$ is the initial value of F, k the relative rate of change, and t is in T. Contrast this infinite richness with a singular qualitative hypothesis such as "All ravens are black". (And yet our theory is about the simplest that can be stated with the help of the infinitesimal calculus.) Nor is a hypothesis of continuity necessary to build theories with infinitely many formulas. The underlying logic itself contains infinitely many tautologies. And any theory that presupposes number theory has not only infinitely many formulas but also infinitely many concepts: ... -2, -1, 0, 1, 2, ... Such infinities are usually manageable thanks to the concept of a set and a suitable notation. Thus instead of listing all the integers, which would be impossible anyway, we introduce the single symbol '$\mathbb{Z}$'. Likewise in defining, say, the infinitely many coefficients of a power series one just writes the *n*th coefficient c_n.

Our growth theory is a factual theory, as shown by the postulates *A*5 and *A*6, which are semantic assumptions and, more particularly, hypotheses concerning relations between concepts in the theory and things and processes outside it. (Actually *A*6 is only a semantic axiom schema, for it specifies neither the referents nor the property P.) In other words, when joined to *A*5 and *A*6, *A*4 is a factual hypothesis and moreover a testable one. To be sure, it cannot be tested directly but only through its consequences *Thm*1 and *Thm*2. On the other hand the axioms of the theory of partial order are not testable: they may be regarded as constituting an axiomatic (hence implicit) definition of the relation "$\leq$". What can be checked is only the conjecture that this or that set satisfies those axioms, i.e. is a model of the theory of partial order. In sum, the difference in reference entails a difference in testability.

Note that our theory of growth is kinematical, not dynamical or mechanismic. Indeed it assumes no growth mechanism. On the other hand, a mechanismic theory of growth would contain statements concerning the mode of growth—e.g. by self-assembly prompted by intermolecular forces, or by cell division, or by sexual reproduction. In such a theory the law statement *A*4 would occur as a theorem, not as a postulate—and possibly in a wider form making room for random fluctuations around the central exponential trend. The demotion of axioms to theorems is a good indicator of increasing depth and therefore power. Thus the famous competitive exclusion principle of ecology, namely "Different species occupy different niches" (or "No two species occupy the same niche"), can be derived from the assumption that the members of different species have different enzymes and, as a consequence, different resource needs and different possibilities of exploiting the environment.

We have just made tacit use of the notion of subtheory, or part of a theory that is itself a theory rather than an arbitrary fragment of it. Thus a mechanismic theory of exponential growth includes, as a subtheory, the kinematical theory of growth. (In general T_1 is a *subtheory* of T_2 if, and only if, T_1 is a theory and T_2 entails T_1 or, equivalently, every formula of T_1

is included in T_2.) The subtheory may or may not have a smaller reference class than the theory of which it is a part: what matters is that all the statements of the former belong in the larger theory but not conversely. Thus particle mechanics is a (tiny) subtheory of body (continuum) mechanics, and the theory of the linear oscillator is in turn a subtheory of particle mechanics.

Subtheories should not be mistaken for *specific theories* or *models* of things of a certain kind. A model contains assumptions that do not occur in the general theory, so it cannot be part of the latter. For example, a specific theory (or model) of a fluid of a certain kind contains not only the general assumptions of mechanics but also the "constitutive relations" or special laws characterizing that fluid. Again, a theory of the evolution of plants is more specific than the general theory of evolution (on which the former is based), for it contains particular assumptions about the roles of chloroplasts and chlorophyll, roots and seeds, etc. So, it is not a subtheory of the theory of evolution. Likewise a theory of sexual selection is not a subtheory of the theory of natural selection, for it contains concepts—such as those of breeding time, competition for mating partner, and courting—that are absent from the general theory. See Figure 9.5.

In general, we may say that a theory T_1 *specializes* to theory T_2, or that T_2 is an *application* of T_1, if and only if there exists a set A of formulas such that the union of T_1 and A entails T_2. The additional assumptions may be further hypotheses (in particular indicators) or data. (Indicators or "operational definitions" are really hypotheses; they may but need not be included in the specializations or applications of a theory but they must occur in every test of it.) For example, we specialize or apply the theory of partial order when enriching it with (semantic) assumptions concerning the nature of R and $\leq$. And we apply particle mechanics when we specify the system to be a pendulum. In all specializations or applications the original reference class shrinks. That is, the special theories embrace logically the

T1
T2
(a)
A
T2
T1
(b)

Fig. 9.5. (*a*) Theory T_1 is a subtheory of theory T_2. (*b*) Specific theory T_2 is based on theory T_1 and contains set A of assumptions not included in T_1.

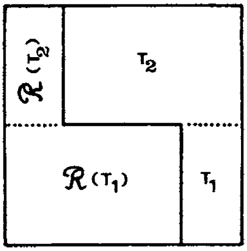


Fig. 9.6. T_2 is a specialization of T_1, and so contains more formulas than T_1 but has a smaller reference class than T_1.

general ones but are referentially included in the latter. I.e. if T_2 is a specialization of application of T_1, then R_2 is included in R_1. See Figure 9.6.

The special theories or models resulting from specialization (or application) can be called *bound models*, for they are attached to general theories. But not all special theories are based in general ones. We call *free models* those that do not result from augmenting the set of assumptions of a theory. (In other words, the totality of formulas of a bound model equals the set of all the consequences of the union of the postulates G of a general theory with the special assumptions A compatible with G, i.e. $P = Cn(G \cup A)$. On the other hand the totality of formulas of a free model is $P = Cn(A)$, for here $G = \varnothing$.)

Whereas in the advanced sciences and technologies most models are bound, in the less advanced ones most of them are free, i.e. they are built from scratch instead of on the basis of general theories. This is particularly the case with the social sciences, where there are no recognized general and valid theories. Needless to say, it takes more imagination to build a free model than to apply a general theory, for in the latter case the existing theory can be used as a scaffolding. Thus no matter how difficult it may be to build a theory of the origin of viri or of fungi, or a theory of the lineage of mammals or of the geographical distribution of plants, the contemporary theoretical biologist can avail himself of general cytological and physiological principles as well as of general theories such as genetics and the theory of evolution. On the other hand the sociologist and the economist have no such guidance. Worse, they are sometimes misguided by theories concerning ghostly entities, such as the free competitive market, or even by ideologies.

The difference between general and specific theories is not always realized or understood. For example, much of the recent work on the foundations and philosophy of classical mechanics takes this to be identical with particle mechanics or reducible to the latter (e.g. Sneed, 1979; Stegmüller, 1976). It ignores continuum mechanics and in particular its constitutive relations (the law statements that characterize kinds of materials). Consequently such studies do not help us understand mechanics or, a fortiori, scientific theories in general. (See Truesdell (1983) for a devastating criticism of such pseudomechanics.) Others criticize the theory of evolution for not making precise predictions, not realizing that extremely general theories, whether in biology or in physics, are not supposed to make predictions unless considerably enriched with special assumptions and data. Thus if we wish to understand, retrodict and predict the evolution of giraffes, we must augument the general theory of evolution with a myriad hypotheses and data concerning the camelids and their presumptive habitats. That is, we must build a (bound) model of giraffes. (For an extremely simple but precise model of evolution see Appendix 1.)

Whether free or bound, the models built in factual science and technology are radically different from those studied by model theory, a branch of logic. (See, e.g. Bell and Slomson, 1969.) The latter are not models of concrete things, such as atomic nuclei and factories, but examples of abstract theories. Thus the natural numbers, together with addition (or multiplication), constitute a model of a semigroup, i.e. they exemplify the concept of a semigroup. Likewise the propositional calculus is a model of Boolean algebra, i.e. propositions satisfy and therefore exemplify the laws of Boolean algebra. This, the *logical* concept of a model, must be sharply distinguished from the *epistemological* concept of a model as a conceptual representation of concrete things of a narrow kind. But, whether logical or epistemological, all such models are conceptual: real things are not models of anything, but instead the objects of modeling. (The statement that the real world is a model of some theory makes sense only in a Platonic ontology.) This terminological remark had to be made in view of the widespread confusion between the two kinds of model introduced by the so-called set-theoretic semantics of science (e.g. Sneed, 1979; Stegmüller, 1976).

We have examined two kinds of inter-theory relation, those of subtheory to theory (e.g. statics to dynamics), and of general theory to specific theory (e.g. group theory to commutative group theory). There are several other inter-theory relations. Let us mention a few of them. *Equivalence*: Two theories have the same composition (set of formulas) but different structures, i.e. they are organized differently—e.g. an undefined notion in one of them is defined in the other, or an axiom in the first is proved in the second. *Example*: the different presentations of the propositional calculus. *Refinement*: One theory constitutes a revision of another—e.g. it adopts a more powerful mathematical formalism, drops certain assumptions and adds others, simplifies some of them and complicates others, etc. *Example*: contemporary classical mechanics *vis-à-vis* classical mechanics. *Reduction*: One theory explains or reduces another—e.g. a black box theory is reduced to, or explained by, a mechanismic one. *Example*: molecular biology explains classical genetics. (We shall study reduction in Ch. 10, Section 3.1.) *Rivalry*: Two theories concerning roughly the same domain are inequivalent because, even though they share some predicates, they relate them differently, hence they explain or predict things differently. *Example*: the information-processing (cognitivist) and the psychobiological theories of vision.

So much for theory structure. We have presented a view that is an elaboration of the standard view adopted by mathematicians and theoretical scientists. This standard view is not accepted by all philosophers. Thus some of them mistake theories for languages, even though a language has no deductive structure and should serve to express mutually incompatible theories; others confuse theories (hypothetico-deductive systems) with hypotheses (single propositions); others try their hand at the alchemical art of dispensing with theoretical concepts in favor of empirical ones; still others, while calling themselves realists, claim that theories are metaphors; finally, others include the applications or specializations of a theory in the theory itself. (See Suppe (Ed.) (1974) for a representative sample of this population.) We shall waste no time with these quaint views for the same reason that we do not take the Flat Earth Society seriously. Those are harmless philosophical myths.

On the other hand we must sound the alarm about a real peril for all theorizing, namely the current computer cult. In fact the overrating of the computer is having extremely harmful effects on mathematics, science and technology alongside its well-known beneficial ones. One is that the general public and the bureaucracy of science feel all the more justified in their contempt for theory as their admiration for The Computer grows. This valuational mistake stems in part from the confusion between theorizing and calculation, and the enormous prestige attained by the latter in connection with space travel and missiles. It is usually overlooked that computation, whether by hand or by computer, involves both theory (e.g.

mechanics) and algorithm (e.g. the infinitesimal calculus), so no computer can replace either. Moreover, when used as calculation auxiliaries computers deliver sheer numbers that make no sense except in the light of some theory. When no theory intervenes in the input or the output of a computer, the operation has no scientific significance. Sometimes it is just a piece of pseudoscience.

A second harmful effect of the computer cult is the growing tendency among investigators to neglect the analysis, improvement, and replacement of existing theories. Indeed, a scientist devoted to computation *uses known theories* instead of examining them or inventing new ones. Not being creative, the computer cannot fill in for the original theorist; and not being critical, it cannot replace the foundations researcher. Because of this emphasis on the application of known theories, what began as the best friend of the theorist is turning against him, and what began as the computer revolution is turning into the computer counter-revolution. (More in Truesdell (1981).)

Thus far we have been concerned only with the structure and content of theories, i.e. we have dealt with the logical and the semantical concepts of a theory. But theories can be viewed in alternative ways as well, in particular methodologically, psychologically, sociologically, and historically. Thus we have at least six different concepts of any factual theory T:

Logical: T is a set of statements,
Sematical: T represents its referents,
Methodological: T helps explain, predict, or design experiments,
Psychological: T is a collection of possible thoughts,
Sociological: T is a product of the social process of inquiry,
Historical: T is a link in a historical process.

Different specialists are likely to emphasize different concepts of theory, and some of them are bound to claim that theirs is the only valid one. In particular, those allergic to mathematics are likely to dismiss the logical concept, and those insensitive to history will ignore the historical concept. However, the six concepts are mutually compatible, not exclusive; moreover, they complement one another. Thus the historian needs to know that a theory is (among other things) a logically organized set of statements rather than an unorganized one, let alone a single statement, for otherwise he will not account for the history of theories proper. And it will not hurt the logician to remember that theories, far from being Platonic ideas, are created, reformed, discarded or forgotten by real people engaged in a historical process. Let us then approach the problem of theory construction.

2. CONSTRUCTION AND REALITY

2.1. *Theory Construction*

How are theories built? What are their building blocks and how are these put together? Are data good theory starters or does one begin with hypotheses? Are there any rules for building theories, so that a machine could be programmed to build them? How about theory reconstruction, in particular axiomatization: are there axiomatization kits? And what is the use of axiomatics anyway? These are some of the problem we shall tackle in the present section.

The building blocks of theories are propositions involving concepts of various degrees of remoteness from ordinary experience. Whereas modest theories are built with ordinary knowledge concepts, ambitious theories are built with either highly refined versions of the former or totally new concepts. Thus the concept of a set refines that of a finite collection, and the concept of a position coordinate that of place. On the other hand the concepts of tautology, vector space, electric field, protein synthesis, and elasticity of demand have no ordinary knowledge antecedents.

Whatever its origin and historical evolution, a concept belonging to a theory may be called *theoretical* if it is either peculiar to the theory or considerably elucidated by it. All of the concepts occurring in logical and mathematical theories are theoretical. And, because all the theories in advanced science and technology contain (usually voluminous) mathematical formalisms (or presuppositions), they contain theoretical concepts galore—in fact infinitely many in the case of any theory presupposing ordinary arithmetic. Some of the concepts of a theory are borrowed from its formal (logical or mathematical) background, whereas others are specific to it. Thus, in the theory of partial order we reviewed in Section 1.2 the concept of implication is taken from logic, whereas the order relation is peculiar to the theory.

The specific or technical concepts of a theory are in turn partitioned into *basic* (or undefined or primitive) and *derived* (or defined). Thus in the theory of partial order " $\leq$" is basic whereas "$<$" is derived. All definitions are intratheoretical statements; more precisely, if explicit they are identities. (Vol. 2, Ch. 10, Section 2.2.) One adds them as needed. Occasionally one needs infinitely many. For example, in elementary number theory every number greater than 0 is defined as the successor of some other number, and so an infinitely long chain of definitions is generated.

The propositions (formulas) in a theory are either premises (postulates), definitions, or consequences of premises or definitions, i.e. theorems. (Recall Section 1.2.) The premises of a formal theory are tentative during the construction period, but cease to be so once the theory has been built and shown to be capable of solving some problems. From here on those premises are stipulations that determine the basic properties of the concepts of the theory: they create them. (Definitions proper, such as "2 equals the successor of 1", are noncreative. But axiomatic definitions, such as that of a partially ordered set, are creative.) The worth of the postulates of a formal theory lies exclusively in their power to create or refine concepts, unify previously scattered formulas, and entail new ones. Any conjectures that may arise in the course of investigating a formal theory are to be proved or disproved. When proved they become incorporated into the theory as theorems, when disproved they are dumped into the dustbin.

Having invented and organized a system of postulates and definitions is important but insufficient to solve any particular problems. To tackle the latter one must enrich that system with subsidiary assumptions. This work of specification or enrichment can be endless: witness the development of group theory, classical mechanics, and quantum mechanics. In all these cases the basic equations are adjoined now one set of subsidiary assumptions, now another, to yield in each case a particular theory capable of handling the given problem or allowing one to pose new problems. Thus in the case of group theory one may add the assumptions that the basic set is finite or else infinite, that the group operation is commutative or anticommutative, and so on. In the case of classical mechanics one may add that the thing of interest is a system of particles or a continuum, that the force is central or noncentral, that the viscosity is low or high, and so forth. And in the case of quantum mechanics one may add that the system is an atom or a molecule, that it is embedded in a magnetic field or an electric field, and so on. In this way intratheoretical progress is made, consisting in constructing special theories (bound models) on the basis of the given general theory. In other cases progress is accomplished by discovering the general theory underlying two or more special and formerly unrelated theories. In still other cases the contribution consists in transforming an untidy theory into an axiomatic system. In all these cases advances are made with only pencil and paper.

The case of factual theories is radically different from that of formal theories. Here every premise that is not a definition is tentative, i.e. a hypothesis not a stipulation. It is supposed to prove its worth, jointly with the other premises, not only by its clarifying, unifying and deductive powers, but also by empirical tests suggesting that it is adequate to facts, i.e. factually true, to some extent. Sometimes the special hypotheses contained in theories are called 'data'—e.g. that a system is composed of so many things of such and such kinds, or that it is confined within such and such boundaries. Actually these are not data but corrigible assumptions which, together with the other premises, sketch the properties of the referent. Data have no place in the foundations of a theory, although every factual theory should make room for them. Thus in the theory of exponential growth in Section 1.2 the initial value $F(0)$ of the property in question, and its relative growth rate k, were left indeterminate, for they are to be determined empirically. Once we get such values we can insert them into the theory and use the latter to calculate, say, the value of F at any desired time t. (I.e. we add to the theory statements of the form "$F(0) = a$" and "$k = b$", where a and b are definite positive numbers.)

There are three reasons for excluding data from the foundations of a theory. One is generality. Thus mechanics is supposed to refer to all possible bodies of a certain kind in all possible (lawful) states of motion. This would not be the case if it contained detailed information about the masses and trajectories of the bodies. The second reason is testability: if a theory contained all the available information it would predict its own evidence and would therefore be untestable. The third reason is predictability: if we could know by sheer experience everything there is to know about things of a kind we would not need theories about them. In sum, the foundations of factual theories do not contain data. They only contain blanks to be filled by data.

Yet there is near consensus among scientists that every factual theory is or ought to be "based on" data such as measurement values. (They have not learned this from their research experience but in the classroom and in textbooks.) But it is not clear what 'based on' means in this context. The consensus view is so ambiguous that it can signify either of the five following theses: factual theories rest on data, they are built out of data, they summarize and extrapolate data, data motivate the construction of theories, or a theory, if true, is supported (confirmed) by data. The last two claims are obviously true. Data, at least those of a new kind, call for explanation, and scientific explanation is best performed with the help of theories. And no theory can be pronounced to be a (sufficiently) true representation of a domain of facts unless it has passed some empirical tests. But the first three claims are false. Let us find out why.

First, as we saw a while ago, data do not belong to the foundation or basis of any theory. This is not a historical accident but a matter of logic. Indeed, every datum is a singular proposition, and singular propositions have no deductive power or, if preferred, they have too much of it, for they point to no hypotheses in particular. More precisely, a datum e implies an arbitrary proposition of the form “If h then e”. (In fact, “$h \Rightarrow e$” is false only in case h is true and e false, which contradicts the assumption that e is true.) Second, most of the interesting properties of things, and all of the essential ones, can be observed or measured only with the help of theories, some of which are involved in the very design of measurement instruments. (For example, a way of measuring the mass of a body is to find its weight and divide the latter by the acceleration of gravity, which in turn is measured with the help of a pendulum and the theory of the pendulum. See Figure 9.7.) Third, no measurement can yield any of the sharp values of any of the

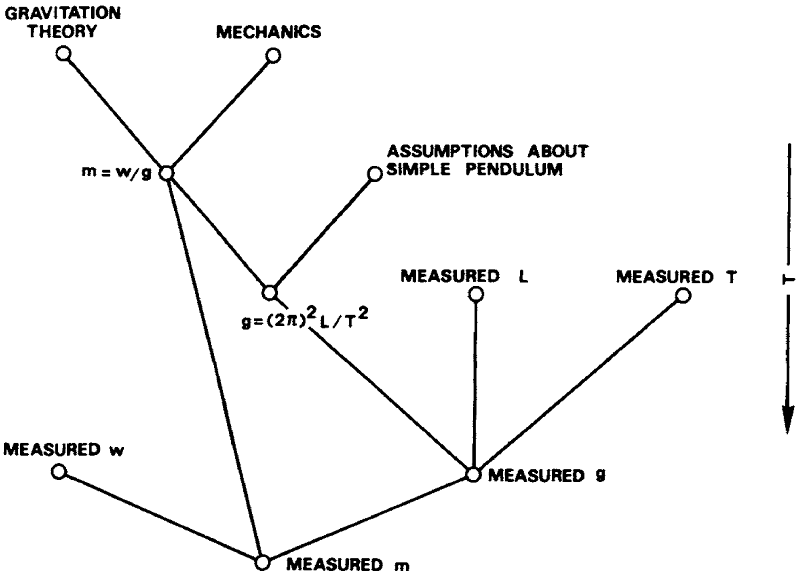


Fig. 9.7. The measurement of mass with the help of a scale (for measuring weight w) and a pendulum (for measuring acceleration of gravity g).

continuous (or piece-wise continuous) functions that occur in all of the advanced factual theories. In particular, the instantaneous values of rates (e.g. speeds) can be calculated but not measured exactly, for measurement yields only averages. (Thus we calculate the rate of change of a property represented by a function F by computing its time derivative, but measurement gives only values of the ratio $\Delta F/\Delta t$ of finite increments.

The problem of whether theories can be built out of data can be approached in the light of the distinction between direct and inverse problems (Ch. 7, Section 4.1). To explain or forecast data with the help of a theory and prior data is a direct problem with a unique (though possibly many-membered) solution. The corresponding inverse problem is: Given a set of data build a theory covering them. Like most inverse problems, this one is indeterminate, both in the sense that data are insufficient, and that there is always more than one theory that explains any given set of data. (Philosophers call this well-known fact "Quine's thesis of indeterminacy of translation".)

Consider the following inverse problem: Given (by observation or by conjecture) the genetic structure of a population at a given time, and assuming a constant environment, find its genetic structure (the distributions of the genes) after a certain number of generations. This is a direct problem that can be solved with the theory of population genetics, which contains precise formulas for simple cases. The inverse problem, of finding such formulas from a study of populations in the wild or in the laboratory, is hopeless, not only because nature does not bother with the simplifying assumptions the theorist is forced to make, but also for the following reason. Most populations in nature have attained a state of equilibrium, i.e. a constant genetic composition, and such equilibrium states can be reached from many different initial states, much as a pebble can reach a given place at the foot of a mountain from many places on the latter. So, the inverse problem is insoluble. Data do not entail theory but may motivate its construction, and theory does not entail data but may help gather them. Only the merger of theory and data can entail further data. Recall Figure 9.7.

If data do not trigger theorizing, what can? Hypotheses can. The theorist usually starts with a handful of more or less disconnected generalizations and, of course, a bag full of mathematical tricks. Some of those generalizations are nothing but empirical generalizations, such as curves that interpolate or extrapolate observational data; others are speculative conjectures. And among the latter some may be mere programmatic conjectures, or schemata to be spelled out, such as "*A* is some function of *B*". These trigger hypotheses may be called the *precursors* of the theory.

Building a system out of precursor hypotheses is not merely a matter of collecting them. They have to be refined and interconnected, possibly by further hypotheses. The refinement may be performed with the help of mathematical tools, and the glue hypotheses may contain concepts not occurring in the precursors. A classical yet often misunderstood example is Newton's creation of classical mechanics. His precursor hypotheses were the kinematical laws conjectured and checked by Galileo, Kepler, Huyghens, and a few others. Newton could not generalize these laws into a general kinematical law because there *is* no such law. What he did was to invent new, dynamical hypotheses containing the new basic concepts of mass and force, to which that of stress was added later on. It turned out that Newton's theory contained not only the precursor hypotheses but also any number of new ones, one for each possible force law, mass distribution, and stress configuration. (Newton's laws of motion are actually schemata that can be specialized by specifying the forces, masses, and stresses.) Likewise the mathematical models that are emerging in neuropsychology contain neurophysiological concepts, such as those of neuronal connectivity and plasticity, and aim at relating and explaining the regularities found by the students of behavior and cognition, as well as many unsuspected ones.

It is difficult to exaggerate the importance of mathematics in the construction of scientific and technological theories. We take mathematics so much for granted nowadays that we sometimes forget the ways in which it helps theory construction. For one thing mathematics offers a huge store of ready made concepts and formulas that, to speak anthropomorphically, are only waiting to be used and interpreted in factual terms. So much so that every factual theory in advanced science or technology is a mathematical formalism together with a factual interpretation of some of its concepts. (See Section 2.2.) For another, mathematics offers a large supply of methods for solving, exactly or approximately, nearly all the problems that are formulated mathematically. So much so that formulating a problem in mathematical terms is usually the hardest stage of a research project in theoretical science, though solving it may be the most time-consuming. (Computers are a great help in solving problems of certain kinds, not in formulating them.) Thirdly, mathematics guarantees exactness and deductive power, and facilitates the gluing and supplementing of the precursor hypotheses of a theory. So much so that the theorist who employs mathematics feels that he is being led by the hand and that his thoughts spread over an entire network that seems not to be his, to the point that he meets unexpected problems and surprising solutions. For a simple mathematical model see Appendix 1.

In the process of building a scientific theory or model with the help of mathematics we can distinguish the following stages:

(i) *survey* of the relevant available knowledge with a vew to finding candidates to precursor hypotheses;
(ii) *invention* of stronger hypotheses, usually containing sophisticated transempirical (nonobservational) concepts, that may entail the precursors;
(iii) *checking* whether the new hypotheses entail the precursor hypotheses or close enough versions of them;
(iv) *mathematization* of the strong hypotheses—often but an aspect of invention, and occasionally requiring the forging of new mathematical ideas;
(v) *investigation of the mathematical problem* with momentary oblivion of the possible factual interpretation—to the point that this stage of research is sometimes taken up by mathematicians;
(vi) *simplifications* facilitating the handling of the mathematical problem or even rendering it tractable;
(vii) *factual interpretation* of some of the results of the previous investigation.

All of these seven steps are equally important. In particular, it is not true that one can always get by without mathematization, or that the essential feature of modeling is simplification. Nonmathematical thinking is clumsy and laborious, inexact and rarely systematic, to the point that purely verbal (i.e. nonmathematical) theories are at best good precursors of mathematical theories. With the help of mathematics we "see" systems, patterns and consequences where none are apparent without it. As for simplification, it is true that it is inherent in theorizing (and also in classification), but it is not more important than invention: the beard must be there before it can be trimmed. And in any case simplicity, if found inconsistent with truth, can be remedied by inventing further concepts or propositions. After all, the historical process has not been that of theoretical simplification but, on the whole, one of progressive complication. Only, such complication is not arbitrary and is not introduced to save some dogma—as is the case with ideological complication—but is motivated by the wish to perfect agreement with fact or a deeper understanding of fact. (See Bunge, 1963.)

In dealing with theories we must avoid several pitfalls. One of them is the very natural attempt to handle new problems with the help of existing theories. While this strategy works in some cases, it is bound to fail in others. Understandably, such conservatism is more likely to succeed in the advanced branches of science and technology. Thus an engineer would be ill advised to try and build a new electrodynamic theory every time he faces a new problem in power transmission or in electromagnetic communication. On the other hand psychologists and philosophers are wrong in insisting that the central nervous system is nothing but a Turing machine, for automata theory does not even contain time, which is essential to account for the variability of neural connections.

A second pitfall to be avoided is that of mathematizing the wrong precursor hypotheses, and embellishing theories that are known to be false or at least are not known to be even approximately true. Contemporary microeconomics, with its powerful mathematics and its slender empirical basis, is a case in point. What is the point of embellishing the neoclassical hypotheses concerning consumer behavior and the way prices are fixed if they are at best untested, at worse at variance with empirical research? Better construct modest econometric models than waste effort studying nonexisting equilibria. (More in Vol. 6.)

A third pitfall is procrastination in starting to theorize for fear that it may be premature. This fear is unjustified, for (*a*) theories, even if false, advance experimental research by suggesting new problems or new ways of approaching known problems; (*b*) knowledge is not properly scientific unless it has a strong theoretical component; (*c*) accumulating empirical information in the absence of theories may prove wasteful, for research that fails to be inspired by theories is bound to yield at best superficial and disconnected data, and at worse irrelevant ones. It is never too soon to start theorizing—and never too late for changing a defective theory for a better one.

Having decided that the time has come to build a theory, can we ask methodology for directions? Is there a logic or technique of theory construction? The answer is in the negative: there are only philosophical fantasies on them. The very idea that original work could be rule-directed is self-contradictory, for originality defies the known rules. (Instead of speculating about the existence of a logic of theory building, we should speculate about the possible neural mechanisms of theory invention, as well as about the social mechanisms that favor or discourage the invention of theories.) Since invention cannot be mechanized, no machine could be programmed to build theories for us. Machines can help once we have created theories and algorithms—e.g. they can help solve some mathematical problems that arise in applying the theories. But even here machines can do only so much. For example, they cannot handle continuity. And their output, namely a bunch of printouts, is far more difficult to interpret than even the least accurate of analytic solutions.

The case of theory reconstruction, or axiomatization, is different: here there is a list of definite do's. (Cf. Bunge, 1973a.) Thus the first step is to investigate the theory to be axiomatized with the aim of spotting the most likely candidates for undefined (primitive) concepts and unproved formulas (postulates)—i.e. the strongest constructs. A second task is to define (implicitly) every undefined notion by means of a set of postulates characterizing both its mathematical form and its factual content if any. (E.g. "F is a piece-wise continuous vector-valued function on the cartesian product of A by B", and "The value of F at $\langle a, b\rangle$, where a is in A and b in B, represents the F-ity of thing a at place b".) Thirdly, we should check whether the proposed system of postulates entails all of the standard theorems in the theory. If not, we should promote some other formula(s) to the axiom rank, or invent an even stronger formula. There are a few other heuristic rules we can use to facilitate axiomatization, but none of them guides the creation of a new theory *ab initio*. In this task only other people, in particular mathematicians, can help the creative theorist.

(It may be objected that modern metamathematics has shown the limits of axiomatizability, particularly with regard to completeness and decidability. For example, the following result has been proved for an arbitrary theory T formalized within the first order predicate calculus with identity: *If T is complete, then T is not axiomatizable iff T is undecidable*: see Tarski *et al.* (1953). Since the vast majority of theories are undecidable, it would seem that axiomatization is the exception rather than the rule. This is not true, for the theorem holds for *complete* theories, and in science and technology there are no complete theories. Moreover, in these fields we have no use for complete theories, as a theory can be applied only if it is *incomplete*, for only then can it be enriched, without contradiction, with subsidiary hypotheses and data. So, the above metamathematical result has no bearing on the axiomatizability of factual theories.)

Should it be asked what the value of axiomatics is, we could answer with another question, namely 'What is the value of cleanliness and tidiness?' In axiomatizing a theory we are bound to discover gaps or inconsistencies, refine concepts, overhaul the mathematical formalism, and reorganize the formulas in a clearer and more reasonable fashion. In the process we may discover new theorems or discover that some formulas that used to be taken for theorems are not really part of the theory. We may facilitate the spotting of the deficient presuppositions or hypotheses to be corrected. We may establish that certain formulas are definitions, and thus save the experimentalist the trouble of trying to test them. We may dispel any doubts concerning the genuine referents of the theory. Also, an axiomatized theory is easier to recall and therefore use. Finally, axiomatization renders a deep and cogent philosophical analysis of theories possible.

Unfortunately a romantic wind is blowing in philosophy once again. Romantics abhor formalization and even analysis, preferring to make sweeping, global, vague, and sometimes derogatory assertions about theories, theorists, and society. There seem to be several causes for this revolt. One is that it is but one aspect of the general movement against science and technology that accompanies both the revival of religious fundamentalism and the youthful rebellion against the establishment. Another is that most of the philosophical formalizations in the past, notably those proposed by Carnap and Reichenbach, turned out to be unfaithful to science and philosophically unilluminating. A third cause is that it is far easier to revolt against formalism than to learn it. This revolt is unreasonable. If one type of formalization does not work, e.g. because it bears on cartoons of scientific theories, or on trivial problems, then one should try a better one. And if one dislikes formalism, one should stay clear of advanced science and technology, which carry mathematics in their very bones.

To conclude, there is no method for building theories other than a command of the basic facts, the founder generalizations, and some mathematical tools. The so-called inductive method does not deliver theories because data collection itself is done in the light of some conjecture or other. Nor does the so-called deductive method yield theories, for it works only where the initial hypotheses have already been conceived. Theory construction is not a rule-directed process but an original brain process to be studied by psychology and social science, not by logic. Such an empirical study of theory construction must pay particular attention to the background knowledge and values of the theorist, to the way he sees the world and our knowledge of it. (See Holton (1973) for the importance of general guiding principles in theorizing.)

2.2. *Theory and Reality*

We have called a theory *formal* if all its referents are constructs, and *factual* otherwise (Section 1.2). Formal theories are then, by definition, unrelated to reality—except of course that they are invented, learned, and used by real beings. On the other hand factual theories are doubly related to reality: semantically and methodologically. In fact they *refer* to certified or putatively real things, and they are *tested* by confrontation with experience, which is a part of reality. In this section we shall be concerned with the semantical relation of representation, leaving the methodological relation of confirmation for Ch. 12, Section 2.1.

Because a factual theory is about certified or putatively real objects, it is more than a formalism: it also includes propositions telling us what refers to what and what represents what. We call these propositions *semantic assumptions*; in the literature they are usually misnamed 'correspondence rules' or even 'operational definitions'. In our view semantic assumptions are corrigible hypotheses, not rules or definitions: we assume them and then check whether the theory containing them passes some empirical tests. If it does, then the system composed of the formalism and the semantic assumptions is pronounced true until further notice. Otherwise we introduce changes in the formalism or the semantic assumption—or abandon the theory altogether. For example, a mathematical model originally designed to account for the propagation of a contagious disease may not do this job but may instead represent adequately the propagation of rumors: the formalism is kept and only the semantic assumptions are modified. See Appendix 2 for an example.

Every semantic assumption is an *interpretation*, i.e. a formula of the form "Concept X stands for (or represents) Y". Some interpretations assign constructs to constructs, others to factual items. We call the former *formal*, the latter *factual* interpretations. Thus the basic (undefined) concepts of the theory of partial order (Section 1.2) can be interpreted as follows:

$$\rho(R) = \textit{The set of natural numbers},$$
$$\rho(\leq) = \textit{Less than or equal to}.$$

This interpretation yields a *model* or example of the abstract theory of partial order: i.e. the formulas of the latter are satisfied under that interpretation. This model does not represent any facts: its semantic assumptions relate concepts to concepts. If, on the other hand, we set

$$\rho(R) = \textit{Collection of lamp posts along a highway},$$
$$\rho(\leq) = \textit{To the left of},$$

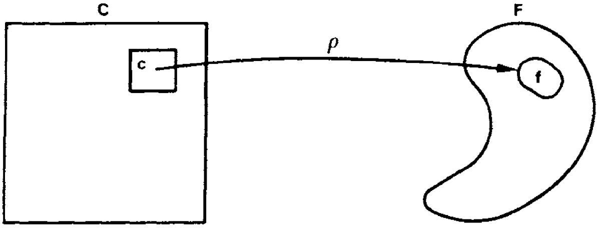


Fig. 9.8 The representation function ρ assigns a subset c of the set C of constructs a subset f of the collection F of factual items.

we produce a model of a collection of factual items. We say that the factual interpretation has produced a *factual model*, i.e. a theory representing a collection of factual items: See Figure 9.8.

The expression 'factual model' is ambiguous, for it signifies an example of an abstract theory and also a model of some factual domain. In the above case the factual interpretation of R and $\leq$ produced a model that is both. Likewise a linguistic interpretation of the theory of semigroups, and an interpretation of Boolean algebra in terms of switching circuits, yield models in both senses. However, not every factual model is of this kind. Most models in science and technology involve mathematical formalisms that have already been interpreted in mathematical terms—e.g. not abstract geometry but Euclidean geometry, not the general theory of functional spaces but the theory of Hilbert spaces. See Figure 9.9. However, the essential point is that factual models, unlike formal models, represent certified or putatively real factual items—material things, their properties, and changes in the latter. Because of this factual reference they must be tested not only conceptually but also empirically.

(Whether conceptual or factual, an interpretation may be construed as a function $\rho: 2^C \rightarrow 2^D$ assigning to every subset of a set C of constructs a subset of a different collection D. If D is taken to be a set of nonabstract but still conceptual items, the interpretation is called *formal*, as when we assign to every subset of C a set of natural numbers, or of points on a sphere, or of real valued functions. Model theory, or the semantics of mathematics, handles only formal interpretations. On the other hand the semantics in Vols. 1 and 2 of this *Treatise* focuses on the case when D is a set of factual items, such as things, properties of things, or events in things. In this case ρ is called a *factual* interpretation, and the corresponding model a *factual* one. Actually factual interpretations are *partial* rather than full functions, for they do not assign to every construct a collection of factual items. Thus a theory presupposing the infinitesimal calculus may contain all the infinitely

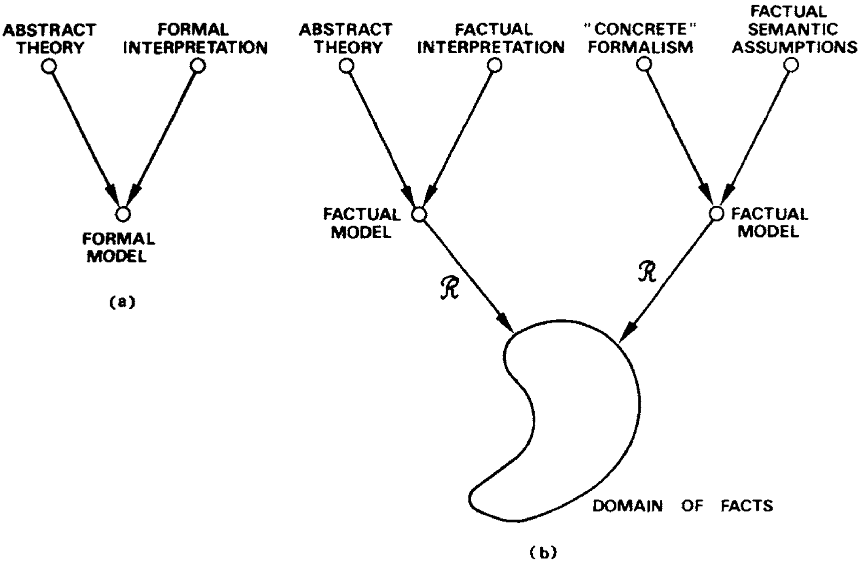

ABSTRACT THEORY
FORMAL INTERPRETATION
FORMAL MODEL
(a)
ABSTRACT THEORY
FACTUAL INTERPRETATION
"CONCRETE" FORMALISM
FACTUAL SEMANTIC ASSUMPTIONS
FACTUAL MODEL
FACTUAL MODEL
$\mathcal{R}$
$\mathcal{R}$
DOMAIN OF FACTS
(b)

Fig. 9.9. Two types of model. (*a*) Conceptual models within formal science: no facts are being represented. (*b*) Conceptual models within factual science and technology: whereas some result from interpreting abstract theories, the great majority result from superimposing factual semantic assumptions on mathematical formalisms that have already been interpreted in mathematical terms—e.g. where sets are not abstract or generic but are composed of specific objects such as real numbers or points in a Euclidean three-space.

many derivatives of every function occurring in the theory, but only the first few derivatives are assigned factual interpretations, e.g. as velocities or accelerations.)

(In view of the preceding we may now characterize theories in a fuller way. An *abstract theory* is, as stated in Section 1.2, merely a logically organized context, and so characterizable as a quadruple $\mathcal{T}_a = \langle P, Q, R, \vdash \rangle$. On the other hand an *interpreted theory* is a logically organized context some components of which are assigned an interpretation, i.e. $\mathcal{T}_i = \langle P, Q, R, \vdash, \rho \rangle$, where $\rho: 2^C \rightarrow 2^R$, and $C = P \cup Q$. In particular, a factually interpreted theory, or *factual theory* for short, is an interpreted theory some components of which are assigned a factual interpretation, i.e. $\mathcal{T}_f = \langle P, Q, R, F, \vdash, \rho \rangle$, where now R includes a nonempty collection F of factual items, and $D = F$, so that $\rho: 2^C \rightarrow 2^F$ is a partial function assigning factual items to some constructs in $C = P \cup Q$.)

The notion of factual interpretation allows us to analyze the set P of propositions of a factual theory $\mathcal{T}_f$. To simplify the discussion, assume that the theory has been axiomatized. Then the set of its postulates is divided into two disjoint subsets: the set M of all *mathematical* axioms, and the set S of all *semantic* assumptions. Every mathematical assumption is a formula of some mathematical theory, and every semantic assumption is of either of the following types:

For sets: "Set A equals such and such kind of (presumably) real entities."

For functions: "Function F represents such and such property of entities of kind K."

A first consequence is that the totality P of propositions of $\mathscr{T}_f$ equals the set of consequences of the union of M and S, i.e. $P = Cn\,(M \cup S)$. A second consequence is that the set P of formulas of $\mathscr{T}_f$ is divided into two disjoint subsets: (*a*) the set P_m of purely mathematical formulas, i.e. those specifying formal properties of the predicates, and (*b*) the set P_f of formulas that are assigned a factual interpretation by virtue of the factual semantic assumptions of the theory. For example,the propositions "A is an algebra of sets" and "F is twice differentiable" are purely mathematical formulas. On the other hand the law statements and the definitions containing factually meaningful concepts, such as the definition of zero centigrade, belong to the set P_f of factually meaningful formulas.

Let us insist that the semantic assumptions are necessary components of all factual theories because a mathematical concept can be interpreted in alternative ways or in none. For example, an increasing function may represent an expanding volume in one theory and a growing population in another. Also, the semantic assumptions are not conventions like the definitions: they are hypotheses and as such they may prove false. For example, in physics an interpretation of an equation as describing a process is said to be 'unphysical' if it allows an event to occur before its cause; and in biology a semantic assumption interpreting a certain function as the total population of an ecosystem is dubbed 'unbiological' if that function takes on negative values.

That the semantic assumptions are hypotheses does not mean that they can be checked empirically in isolation from the formulas they are attached to. What are testable are these formulas together with their factual interpretation or, rather, the factually interpreted formulas. For example, the formula "$y = ax + 1$", where every symbol designates a positive real number, is not empirically testable. On the other hand the following is: "$y = ax + 1$, where 'y' represents the ratio of lengths of a metal bar at temperatures x and 0, and a is the expansion coefficient of that particular metal".

Any given thing can be represented by different theories. This is not a datum but an epistemological hypothesis confirmed but not proved by historical research. It is an optimistic hypothesis that encourages theorizing and moreover it is an irrefutable hypothesis. Indeed, were we to face things of some kind that appeared to resist theoretical modeling we would rather blame our own incompetence than the things themselves, and we would hope that others should succeed where we have failed.

Factual theories represent their referents but are not perfect pictures of them. They are not pictures at all but symbolic constructions bearing no resemblance to the objects they represent. (Boltzmann wrote that the relation of theories to reality is like that of letters to spoken words, and Einstein that it resembles the relation between numerals and coats in the cloak room.) Nor can theoretical representations be perfectly faithful: they are at best approximately true. Hence naive realism is just as wrong as conventionalism. Critical realism is the ticket.

The different theories representing things of a given kind are not necessarily equivalent: some may be deeper or truer than others. For example, consider the following sequence of models of the terrestrial atmosphere that have been considered historically, and are still being studied:

1. Flat earth, homogeneous film of dry air, no radiation.
2. Round earth, homogeneous film of dry air, no radiation.
3. Flat earth, different strata of dry air, no radiation.
4. Round earth, same.
5. Flat earth, strata of moist air, no radiation.
6. Round earth, same.
7. Flat earth, each stratum of moist air divided into homogeneous parallelepipeds, no radiation.
8. Round earth, same.
9. Flat earth, same plus radiation.
10. Round earth, same.

Each model is more realistic than the preceding in at least one respect, and the last is the most realistic (the truest) of all. But even the first model, though very coarse, has some use, for it allows for winds—although, without radiation, it cannot explain them, and without moisture, it makes no room for precipitation.

This concludes our accout of the semantics of factual theories. (For a more detailed treatment see Vols. 1 and 2.) Our account differs radically from the view commonly held by mathematicians or philosophers, which neglects the factual semantic assumptions. One such popular view is *Pythagoreanism*, according to which things *are* constructs instead of being represented by constructs. Thus some writers state that a particle is an oriented one-manifold, an electromagnetic field a tensor field, and in general a physical system "a set of individuals together with numerical functions on these individuals" (Sneed, 1979). This view is incapable of distinguishing between mathematics and factual science. It does not explain how an unchanging construct, such as a tensor field, can be identical to a changing thing, such as an electromagnetic wave. Besides, it is contradictory, for it identifies a given construct now with one thing, now with another, instead of stating that the construct represents one thing in one theory, and another thing in another.

A related view is *formalism*, according to which model theory, i.e. the semantics of mathematics, suffices to analyze factual theories (Sneed, 1979; Stegmüller, 1976). The upholders of this view have probably been misled by the ambiguity of the term 'model', which in metamathematics designates an example of an abstract theory, whereas in science and technology it designates a specific theory (Section 1.2). Consequently they ignore the concepts of reference and representation, central to the semantics of factual theories, and they cannot explain why factual theories cannot be validated or invalidated by purely mathematical considerations. (For further criticisms of Pythagoreanism and formalism see Vol. 2, Ch. 6, Section 3.)

Finally there is *operationism*, or the doctrine that factual theories are mathematical formalisms enriched with operational "definitions", i.e. descriptions of the way the properties represented by the theory are measured. This view is far closer to the truth than the two previous views, but it confuses matters of reference and representation with questions of test. It ignores that one and the same property may, in principle, be measured by different methods. (For further criticisms of operationism see Vol. 1, Ch. 5, Section 3.4, and Bunge, 1973a, Ch. 1.)

In contrast to the above views ours is realist and it has been incorporated into the axiomatization of several basic physical theories (Bunge, 1967b).

3. CONVENTION AND LAW

3.1. *Convention*

There are several kinds of convention, and every theory contains at least conventions on nomenclature. Some conventions, like counting the negative numbers on the left side of an axis, are mere customs (that might not have been adopted if we all were left-handed). Others, like definitions, are time-saving and clarity-gaining devices. Some conventions are dispensable and all are replaceable. Also, some conventions are more useful than others. For example, in daily life the meter is handier than either the fermi, convenient in atomic physics, or the light year, convenient in astronomy.

We shall study conventions of the following kinds: notational, definitions, conventions about scales, and simplifying assumptions. *Notational conventions* assign nonlinguistic objects to signs. They are of two kinds: designation rules and denotation rules (Vol. 1, Ch. 1). A *designation rule*, such as "'$\mathbb{N}$' designates the set of natural numbers", assigns a construct to a symbol. (Nominalists do not need designation rules and they have no concept and no name for that which remains invariant under changes of name.) And a *denotation rule*, such as "'W' denotes the direction in which sunset occurs", assigns a factual item to a symbol. Similar conventions are those concerning the positive direction of rotations, the sign of electric charges, and the choice of the initial and final stage of a process. Every single notational convention is replaceable, though perhaps at the risk of confusion. But every explicit context, in particular every theory, should contain a set of notational conventions—unless they are notorious.

A *definition* is a horse of a different color: it stipulates the identity of two concepts or of their symbols. Thus we may define "positron" as "positive electron" and "1" as the successor of 0 in the set of natural numbers. Such identities are adopted by convention, whereas others, such as "For all real numbers $x, (-x)^2 = x^2$", are provable. The specific difference between a definition and an identity is then that the former is stipulated. Such difference is strictly methodological. On the other hand the identities "Heat = Kinetic energy of randomly moving atoms or molecules", "Gene = Segment of DNA molecule", and "Mental state = State of a plastic neural system in a living brain", are hypotheses to be justified empirically. They state identities of factual items, not of concepts; they state the synonymy of different expressions.

Whether or not a formula is a definition depends on the context in which it occurs. Thus whereas in elementary mechanics forces are undefined and linear momenta defined, in Hamiltonian dynamics forces are defined and linear momenta are primitive. And depending on the context we may postulate or prove the existence of other minds. (In principle it can be proved from the basic similarity of all human brains.) The contextuality of definitions is just a particular case of the contextuality of concepts. Thus the concept of mammal does not occur in pure mathematics, and that of creation out of nothing is out of every scientific context. Only mathematical and philosophical concepts can cross all disciplinary borders.

That definitions are conventions does not mean that they are wholly arbitrary. In a well-organized context the place of a definition cannot be shifted without reorganizing the entire system, and the defined concepts cannot exchange places with the defining ones. For example, numbers can be defined in terms of sets but not conversely. Likewise, "process" can be defined in terms of "thing" but—*pace* Whitehead and his followers—the converse is impossible. (Indeed the very notion of a process, or sequence of states in a thing, involves some concept of a thing. Therefore the definition of "thing" as "a bundle of processes" is circular.) Given a choice of undefined concepts, the strongest ones are preferable, and the defined ones cannot be introduced haphazardly but in an orderly fashion. (Thus the definition of state of a thing presupposes the concepts of thing and of property of a thing.) Moreover, definitions are sometimes backed by axioms or theorems, so they cannot be introduced until such formulas have been stated. Thus the notion of an inertial reference frame is definable in terms of equations of motion, or of field equations, which in turn presuppose the general concept of a reference frame, as well as the concepts of space, time, and much else.

Definitions relate concepts but this does not mean that they lack factual reference or factual presuppositions. For example, the definition of "100° C" as the (same as) the boiling point of pure water at sea level has a factual content: it refers to water. And the definition of "machine" as "artifact for transforming energy of some kind to some purpose" presupposes the law of conservation of energy. Indeed a device alleged to create or annihilate energy would not be called a machine.

Even if a definition has an external reference, it does not introduce new knowledge and cannot take the place of a hypothesis. (This does not hold for axiomatic definitions, which are creative and, in the case of science, are systems of hypotheses.) In particular, a definition belonging to a theory in pure mathematics does not describe any processes in the real world, not even if interpreted in factual terms. (This is ignored by the Bayesian school of epistemology, according to which Bayes' theorem represents the way we learn from experience. This is false, for the theorem is an immediate consequence of the definition of conditional probability.)

Since definitions define only concepts, they cannot define things. This must be stressed because some philosophers believe that the referents of factual theories can be defined the same way concepts are defined, e.g. in terms of a system of equations. What such a system does is to *describe* its referents not to define them. What is true is that a theory may help *identify* factual items as members of a class. For example, Newton's mechanics, together with certain measurements, helps us identify bodies, genetics genes, and so on. In short, an adequate (sufficiently true) theory (*a*) describes its referents, (*b*) defines some of its own concepts, and (*c*) is a criterion for identifying (not defining) its referents. On the other hand most definitions are not utilizable as criteria. Thus a definition of consistency is no criterion for proving or disproving the consistency of a theory, and a definition of truth does not suffice to assign truth values.

All factual theories, and all measurements, involve *units*: of length, time, mass, etc. Units are conventional and they are involved in the very characterization of magnitudes such as physical quantities. Thus when giving the value of the mass of a body b relative to a reference frame f, it is essential to mention the unit u in which it has been measured or computed. In short, such a datum or assumption has the form "$M(b,f,u) = m$", where m is a positive real number. (In general, the property mass is represented by a function of the form $M: B \times F \times U_M \to \mathbb{R}^+$, where B is the set of all bodies, F that of all reference frames, and U_M that of all conceivable mass units, whereas $\mathbb{R}^+$ is the set of positive real numbers. See Bunge (1971).)

The meter, the second and the kilogram can be taken as fundamental units; the micron, the year and the gram as derived units. Units are conventional but not wholly arbitrary. Thus the second, or more precisely the mean solar second, is defined as 1/86,400 mean solar day, which in turn results from astronomical observations and calculations. The fundamental units are materialized as standards, which are reproducible in any well equipped laboratory. And such standards, far from being arbitrary, are chosen because of their accuracy, stability and reproducibility. A good example is the atomic length standard, defined as the wavelength of a certain sharp spectral line that is easy to prepare and measure, and that has a known theoretical description (namely as the result of a certain atomic transition).

A full homogeneous scale is equipped not only with a unit but also with a starting point or *zero*. The absolute or Kelvin zero of temperature is one example. The year 1 A.D. is another such conventional point of departure—the number 0 having been unknown to the makers of the Western calendar. Such choices may look natural and in a way they are, but they are conventional. Thus we could shift the zero of temperature below its current position by an arbitrary amount. However, the additional stretch would have no use, since atomic and molecular motion cease at the universally accepted absolute zero. (As a consequence we would have to subtract a certain constant from all the energy values.)

Next in our list of conventions come *simplifying assumptions*. All except fundamental theories contain some simplifying assumptions. For example, in a certain calculation or measurement we may pretend that the value of *pi* is 3.14; that the system of interest has no internal structure, or is perfectly well isolated, or has only the properties we happen to know; that certain discontinuous variables, such as population, are continuous—or conversely; that brain functions are not affected by social circumstances; that unemployment or inflation has a single cause, etc. All such simplifying assumptions are introduced to expedite modeling or inference, or even to make them possible. They are white lies, for they serve the search for truth. They become black lies when presented as truths or when we keep them even after having been shown to be wide of the mark—as is nowadays the case with the assumption of perfect competition, still at the heart of most economic models.

In sum, all theories contain conventions and, paradoxically, factual theories contain the most. The epistemological moral is obvious: Although factual theories may represent their referents in a fairly true manner, such representations are not pictures or copies.

3.2. *Law*

Man is a compulsive pattern seeker. We abhor the idea of chaos and tend to see regularity even where there is none. Thus, unless we are familiar with probability, we may feel tempted to commit the gambler's fallacy, i.e. to expect that a streak of bad luck will be compensated for by one of good luck, and conversely. We expect order and this is why we look for it. But order is seldom perceptible: the phenomenal world, unlike the real world, is chaotic. Therefore pattern must be either hypothesized or imposed. We hypothesize pattern when we hazard a hypothesis about some constancy or some trend; we impose order when we enforce a rule of action, i.e. compel things to fall into step.

Let us examine the ideas of pattern, trend, empirical regularity, law, and rule. Any stable or reproducible configuration of objects of some kind may be called a *pattern* even if it is the outcome of chance. *Examples*: a string of pebbles left on the sand by the tide, a finite sequence of tones, a behavior pattern. Common to all these instances are order and stability (or repeatability or reproducibility). Every pattern is some order in some collection of objects, and can thus be represented as a structured set. The structure is a set of relations, possibly functions. The converse if false: even a totally chaotic (but finite) set of points on a plane can be represented by a function. Only, such function will not belong to a family, will not be continuous, will not solve an equation, etc.: it will be an *ad hoc* or stray function. More on this below.

Patterns are stable for a while but need not be eternal. Internal tensions among the members of a system, or environmental changes, may end up in pattern changes. Thus a teacher, trainer or experimenter can teach an animal to adopt new behavior patterns by rewarding actions of certain kinds. But he can just as easily extinguish the newly acquired behavior pattern by negative reinforcement. In either case the teacher forces a pattern change, or repatterning. (He does so with the help of certain learning laws.) In the case of biosystems and sociosystems repatterning can occur without change in species. Thus a biopopulation occupying a new territory, and doing well in it, does not change species immediately. Likewise a social system surviving a drastic change, such as revolution or war, may still be what it used to be, e.g. a production center or a school. On the other hand, in the case of physical and chemical systems repatterning is always accompanied by changes in kind or species.

One of the most interesting problems is that of the emergence of pattern from the interplay of initially independent, hence randomly behaving, entities. For example, dust is distributed irregularly throughout the universe, but gravitational attraction clumps it; if several penduli are suspended from an elastic bar, they end up by oscillating synchronically regardless of their initial positions; any piece of cast metal is a random agglomeration of grains of irregular bulks and shapes, but when drawn into a wire or rolled out into a sheet they tend to adopt a common orientation; genic mutations and recombinations are random, but the environment makes certain uniform demands and thus produces some order; likewise the actions of persons are constrained by nature and society, and thus fall into patterns. In all such cases the scientific problem is to explain the emergence of pattern, or the repatterning, in terms of laws and circumstances. And the technological problem is to favor, channel, or prevent processes of pattern emergence and repatterning.

A pattern is called a *trend* when it is a uniform sequence of states of, or events in, a system. Thus the expansion of the universe, and the increase in the prices of most commodities over the past decade, are trends. So are the increasing effectiveness of the hands along the evolution of primates, the steady increase of the total human population since the industrial revolution, the increasing literacy and the rising life expectancy in the Third World, and the increasing lethality of weapons.

Trends are stable while they last. They can stop or even reverse, particularly if the events concerned are brought under human control. Thus at the time of writing it is conceivable that, at some point in time, the universe will start contracting and prices will start falling. The uncovering of trends is an important task in every discipline. It is often the work of educated intuition, and sometimes the result of applying certain statistical techniques. (The whole of econometrics is an effort to discover and represent economic trends.) But we should not rest content with knowing about trends: we should try and explain them in terms of laws and circumstances.

Trends are not laws but they are often confused with laws in the life and social sciences. A good example is the so-called Lartet law, according to which the brains of mammals and birds have increased steadily in relative size along evolution. This is a typical trend, so much so that most of the species to which it applies have become extinct. Or take Marx's objective of discovering what he called "the laws of motion of the capitalist society". He did not attain it, but he did discover a number of trends or short term tendencies. Some of them, such as the increasing concentration of capital, are still in force; others, such as the increasing pauperization of the working class, have been reversed in the industrialized countries. (More on the difference between trends and laws in Popper (1957) and Bunge (1967a).)

Next in line come the empirical generalizations occurring in all branches of factual knowledge. A formula may be called an *empirical generalization* if it is a summary of data or an inductive generalization from data. In science and technology the typical empirical generalization is the smooth curve fitting a finite set of observational or experimental data: Figure 9.10. The empirical generalizations of this type are full fledged hypotheses, not data, for they interpolate between data: they concern not only actuals but also possibles. Consequently they are less, not more certain than the data they cover—so much so that every set of data, being finite, can be covered by infinitely many smooth curves passing near them. True, in the absence of a theory we usually choose the simplest such curves. However, simplicity is convention, not law. Indeed further research may well, and often does,

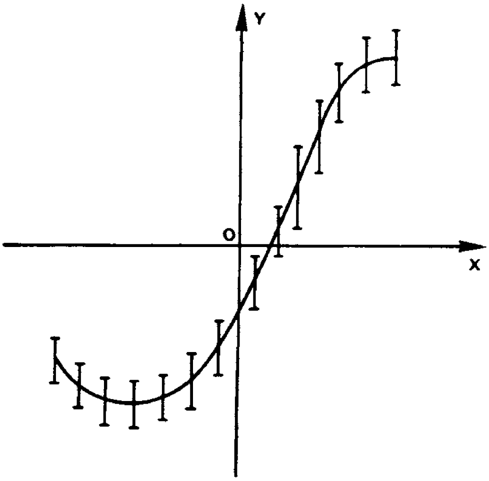


Fig. 9.10. One of the infinitely many possible curves passing through a set of data. The vertical segments, possibly unequal, represent the errors of observation.

refute the simplest hypothesis, forcing us to adopt a more complicated one exhibiting the fine structure of the factual domain in question. See Figure 9.11.

An empirical generalization is supported only by the data that suggested it: it is a stray not a systemic hypothesis. Should new unfavorable data appear, we will be forced to give it up, for we have nothing else to fall back on. If the generalization has been obtained by using a standard in-

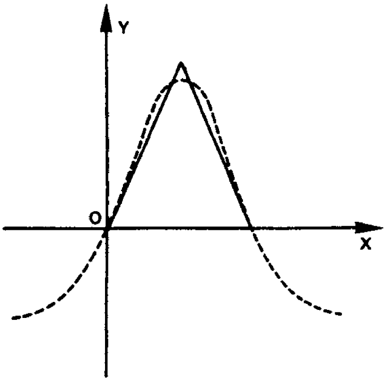


Fig. 9.11 Data suggest full line. Theoretical investigation suggests dotted line.

terpolation or curve fitting technique, it can easily be corrected by changing the values of some of the parameters occurring in it. But this methodological advantage has a high semantic price, for such parameters have seldom a factual meaning. So, paradoxically, empirical generalizations involve concepts without real counterparts. As we shall see in a moment, this is in sharp contrast with fundamental laws.

Laws are patterns of a very special kind: they are systemic, not stray, and they do not change unless the things that satisfy or "obey" them are metamorphosed, i.e. change into things of a different kind. The laws of constructs are propositions, such as the associative and the commutative laws for the addition of numbers. On the other hand the laws of concrete or material things are properties of the latter—and not just any properties but their essential properties. These objective patterns, like the trends, are representable by propositions. We call them *law statements*. This difference, nonexistent in an idealist epistemology but central to a realistic one, is a particular case of the fact-construct difference. In every case a given factual item, be it state, event or pattern, is representable by alternative propositions, some truer than others.

A law statement is a hypothesis asserting that every thing of some kind has a certain property. The following are typical, though by no means exhaustive, forms of law statements found in science and technology. (*a*) "If x is a thing of kind K, and x possesses property P at any given time, then x conserves P at any other time". (*b*) "Every thing possessing property P possesses also property Q". (*c*) "All events of kind E have the same probability p". (*d*) "If x is a thing of kind A, and x undergoes a change of type P, x becomes a thing of kind B". (*e*) "The probability distribution of property P of things of kind K changes in the course of time according to such and such formula". (*f*) "Properties F and G of every thing of kind K are lawfully related to one another". (The mathematical form of this programmatic hypothesis is as follows. Let $f: A \rightarrow B$ and $g: A \rightarrow B$ be functions representing the properties F and G respectively. Then F and G are *lawfully related* if, and only if, there is a third function $h: A \rightarrow B$, representing another property of things of the same kind, such that g is the composition of h and f, i.e. $g = h \circ f$. See Figure 9.12.)

Law statements have occasionally been construed as prohibitions (Popper, 1959). Actually law statements tell us what is possible, from which we infer what is not. What happens is that, in certain popular versions, a handful of law statements are presented as statements to the effect that certain facts are impossible. Here are three well-known examples. (*a*) The

$$\begin{tikzcd}
a \arrow[r, mapsto, "f"] \arrow[dr, mapsto, "g = h \cdot f"'] & f(a) \arrow[d, mapsto, "h"] \\
 & g(a)
\end{tikzcd}$$

Fig. 9.12. Functions f and g represent properties of thing a. They are lawfully related if there is a third function h, which also represents a property of a, and such that its composition with f equals g.

principle of conservation of energy is sometimes translated into pragmatic terms (i.e. transformed into a nomopragmatic statement) for technological purposes, namely thus: "It is impossible to construct a perpetual motion machine of the first kind". But the original principle is affirmative, nonmodal, and refers to closed systems in general, not necessarily machines. (*b*) Pauli's exclusion principle states that the different components of a microsystem are in different states, or that the probability of their being in the same state is zero. But it can be reformulated in a more intuitive way, as "No two components of the same microsystem can be in the same state". (*c*) The competitive exclusion principle of ecology was originally formulated as "No two species can occupy exactly the same niche", or "Complete competitors cannot exist". A more correct (and more easily formalizable) formulation is: "Every species has its own peculiar niche" (or "There is a bijection between the set of species and the set of niches"). In short, law statements are not prohibitions but they can be used as premises for deriving prohibitions. By delimiting the set of lawfully possible events we *eo ipso* characterize the set of conceivable but factually impossible events.

The specific differences between a law statement and any other factual statement are generality, systemicity, and confirmation. More precisely, we adopt the following definition (Bunge, 1967a). A factual statement in a theory is a *law statement* if, and only if, (*a*) it is general in some respect (i.e. refers to every single member of a species, or extends over an entire region of spacetime), and (*b*) it has been satisfactorily confirmed (for the time being). New data may restrict the domain of truth of a law statement. They may show that the latter holds only between certain bounds of certain variables; or, on the contrary, they may show that the law statement holds beyond the domain originally postulated. And the data may also refute the statement, in which case it is demoted from the rank of law statement, and the whole theory containing it is put in jeopardy.

This suffices to show how different laws are from definitions, with which they are sometimes confused. For example, Mach and his followers regarded Newton's second law of motion as a definition of "force". This is a mistake, for (*a*) there are a number of distinct force laws, which must be hypothesized separately from the law of motion, and (*b*) the latter, in conjunction with a definite force law, is empirically testable—and moreover it has been found false in the very small as well as in the very large. (For a criticism of Mach's attempt to define "force" and "mass" see Bunge (1966).)

The most precise representation of a law is with the help of functions. But not every function qualifies. Thus the function f that maps the set $\{a, b, c\}$ into the set $\{d, e\}$ in such a way that $f(a) = d$, $f(b) = e$, and $f(c) = d$, is a stray or *ad hoc* function that represents no intrinsic structure of its domain. We shall call a function *systemic* if it is the object of a theory and belongs to a system or family of functions, i.e. a set whose members are mutually related in precise ways—e.g. by recursion formulas. Otherwise the function will be said to be *stray*. For example, the trigonometric, spherical, cylindrical and hypergeometric functions form systems. The concept of a systemic function allows us to define lawfulness in this way: A set of factual items (states, events or processes) is *lawful* if it is representable by, or with the help of, systemic functions.

Probabilities are systemic for they are subject to the theory of probability. They occur in all *probabilistic* (or *stochastic*) law statements. Notwithstanding the opinion of distinguished philosophers (notably Nagel (1961) and Hempel (1962)) probabilistic law statements are universal in form and scope. For example a velocity distribution law states that, for *all* systems of a certain kind (e.g. molecules in a gas, or cars in dense traffic), the probability of an individual component having a velocity comprised between v and $v + \Delta v$ equals $p(v). \Delta v$, where p is the probability distribution characteristic of the system. On the other hand statistical generalizations, which are empirical, are not universal. In fact, a typical statistical generalization is of the type "f percent of A's are B's". Every such generalization concerns a population as a whole, whereas the corresponding probabilistic laws concern every individual of a given population.

Randomness is not the opposite of lawfulness but one type of it. Thus every outcome of a coin flipping experiment has a definite probability. A sequence of events can be said to be *random* if every event in it has a definite probability; otherwise the sequence is either chaotic or causal. (This is not the standard definition of randomness—but then there is no standard definition.) Randomness can exist objectively, as in the case of atomic and molecular events and of genic mutations; or it can be manufactured, as in the case of randomization and random sampling. Interestingly, random sampling can be used to uncover order, not only in factual science and technology but also in mathematics. This is the case with the Monte Carlo method for solving certain numerical problems. (*Example*: Suppose the problem is to compute the area of a region E of the plane included in the unit square. To do so select at random n points from the unit square, determine their positions x_i, and find out whether they belong to the given region E. That is, assign a value to the characteristic function χ_E of the set E to each sampled point. The Monte Carlo approximation is: Area of $E \cong (1/n) \sum_{i=1}^{n} \chi_E(x_i)$).

Laws, whether or not probabilistic, should not be mistaken for processes, in particular trajectories, life histories, and histories. Thus it is false that every one of us has his own laws: we all share a number of laws but differ in our circumstances, hence in our life histories. Every process can be analyzed into laws and circumstances. And every set of laws is compatible with many different processes, as many as different sets of circumstances. The particular circumstances are representable in the law statements by assigning special values to the parameters occurring in them or in their logical consequences. Thus there is presumably a single law of biological growth, but every pair $\langle$species, environment$\rangle$ is characterized by its own parameters. In other words, a single law covers an entire bundle of trajectories or histories. See Figure 9.13.

Given laws and data we can infer (tentatively) the corresponding processes. The inverse problem, to infer laws from descriptions of processes, is logically impossible because every law encompasses, in principle, infinitely many possible processes. In other words, law statements cannot be inferred from a knowledge of processes or histories. Hence they must be hypothesized—and checked. Far from being inferred, law statements are inference tools. For example, a recursion formula may allow one to infer the nth member of a family of functions from a knowledge of the $(n-1)$th or the $(n+1)$th or others. (Thus, the discrete analogue of the exponential function is a solution of the recursion formula $f_{n+1} = af_n$, where a is a parameter and n a natural number.) A law statement containing a time variable helps one predict the time at which the thing in

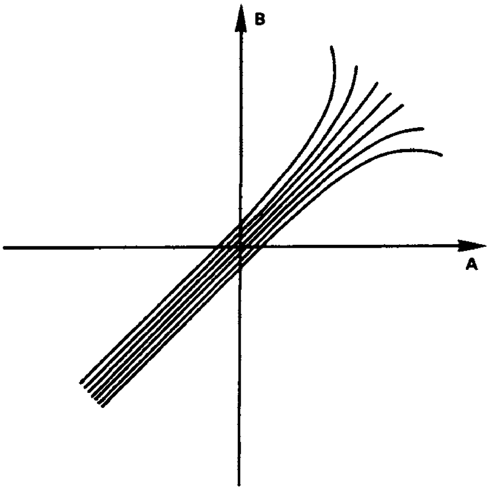


Fig. 9.13. A whole bundle of trajectories (e.g. life histories) in state space is described by a single law. Equivalently: a law is a bundle of possible histories.

question reaches a certain state. Likewise a law statement containing spatial coordinates helps us infer what happens at a given place from a knowledge of what is or was happening at another (e.g. on a surface surrounding the given place).

It is instructive to compare law statements with empirical generalizations, the former being the rationalist, the latter the empiricist paragon of scientific construct. Firstly, the empirical generalizations of the quantitative type, obtainable by interpolation techniques, are usually polynomials, which are stray not systemic functions. Not so the law statements, which contain systemic functions. Secondly, such polynomials contain adjustable parameters (one more than the degree of the polynomial). On the other hand the law statements, particularly of the mechanismic type, have a minimum of adjustable parameters. (The basic equations of electromagnetism and mechanics, whether classical or quantal, have no adjustable parameters at all: all the constants they contain are universal.) Thirdly, unlike the adjustable parameters in the empirical generalizations, the constants in the basic law statements have some factual meaning or other: they are universal constants. Fourthly, most basic law statements are mathematically more complicated than polynomials; even simple laws such as those of the oscillation period of a simple pendulum, and Snell's law of refraction, go far beyond polynomials.

Because of all these differences between law statements and empirical generalizations, the empiricist epistemology, which favors the latter and mistrusts or even rejects the former, does not fit the facts of scientific practice. Nor does critical rationalism, for which all hypotheses are groundless, none being better than any others except that some resist better the attempts at refuting them (Popper, 1959, 1963, 1974). As we saw before, a law statement is not just any hypothesis: it is a well-confirmed systemic hypothesis, hence one supported not only by data of a certain kind but also by all the data that support all the other hypotheses in the theory. Thus the laws of the hydrogen atom and the hydrogen molecule, of radiative decay and radioactive decay, of semiconductors and superconductors, derive from the basic laws of the quantum theory in conjunction with special assumptions; so, every time one of them is confirmed or refuted, its companions are confirmed or refuted.

Law statements can be classed in many ways: according to their logical form, mathematical properties, the kind of thing they refer to, their deductive power, generality, etc. (See Bunge (1963) for some such classings.) Here we shall mention only two distinctions, those between basic and derived, and between general and special. A *basic* law statement is one that occurs as a postulate or axiom in a general theory. Any logical consequence of basic law statements and definitions will be called a *derived* law statement. For example, the laws of conservation (of energy, linear momentum, etc.) follow from the equations of motion or field equations, which in turn follow from variational principles.

Our second partition of the set of law statements is into general and special. Since these terms are relative we must qualify them. We shall say that a law statement is *universal* in the collection F of factual items if it refers to every member of F—whether it holds for every one of them being another matter. The statement is *particular* in F if it refers to every member of a proper subset of F. The utmost general law statements are those concerning things of all kinds, from elementary particles to societies: they are the *universal* laws of ontology, such as "Every thing changes", "Nothing comes out of nothing", and "Every event is lawful". Then come the *generic* laws, i.e. those referring to all things of a given genus, e.g. bodies, electromagnetic fields, quantons, cells, societies, etc. Finally come the *specific* laws, i.e. those concerning all the members of a given species of thing, such as the so-called constitutive relations of continuum mechanics and electrodynamics, or all vertebrates, or all capitalist societies. Such specific law statements occur in all models or specific theories. Clearly, the reference class of a specific law is included in that of the generic laws, which in turn are included in the reference class of the universal laws. And even the most specific of all laws, such as those concerning the motions of individual planets and their satellites, are general in some respect, e.g. they hold for all times. In other words, scientific laws, just like the conventions of social behavior, are not concerned with minutiae: *De minimis lex non curat*.

Our last type of pattern is the rule. Rules are precepts or recipes for doing something. Unlike law statements, rules do not describe, explain or forecast: they prescribe. (The difference between rules and factual laws can be elucidated with the help of the state space formalism used in Vols. 3 and 4. A *law* of things of kind K is a mapping from the state space S_K of the K's into itself, i.e. $L_K : S_K \rightarrow S_K$. Thus, if s is a state of a thing of kind K, i.e. s is in S_K, then "$L_K(s) = t$", where $t \in S_K$ represents the lawful passing of the thing from state s to state t. On the other hand a *rule* may be construed as a mapping from the set of all possible states of the system of interest to a set of possible actions of the agents dealing with it. More precisely, a rule prescribing the behavior of animals of kind M with respect to things of kind K is a mapping of the state space S_K into a collection A_M of actions in the behavioral repertoire of the individuals of kind M. I.e., $R_{KM} : S_K \rightarrow A_M$. Thus if s is in S_K, then "$R_{KM}(s) = a$", where $a \in A_M$ is the action prescribed by rule R_{KM} for individuals of kind M facing a system of kind K in state s. The distinction between rule and law disappears only when $M = K$ and $A_M = S_K$, i.e. when agents and patients are the same, which is the case of the rules of social action. We shall return to this important point in Vol. 6.)

An *instruction* can be construed as a set or a sequence of rules for handling problems of some kind. If the problem is conceptual, the instruction directs thinking; if practical, it controls doing. An *algorithm* is a special kind of instruction, namely a sequence of rules for performing effective operations on symbols, and sure to halt at the result that is being sought. For example, an instruction for extracting square roots, or for computing the values of some function, is an algorithm; a correct computer program too can be regarded as an algorithm. Algorithms are supposed to be consistent with laws, but they are not laws themselves; sometimes they are just clever law-abiding tricks. In particular, as usually understood the laws of mathematics are about concepts, whereas algorithms are rules for manipulating symbols designating concepts. Besides, one and the same set of laws is consistent with a variety of algorithms, some more efficient than others. Thirdly, whereas all the mathematical laws are parts of theories, algorithms can be *ad hoc* and in any case they do not constitute hypothetico-deductive systems. So much for the algorithmic view of mathematics.

Animals and machines of certain kinds can be programmed or taught to follow instructions of certain types. Thus a Turing machine (an ideal computer) is characterized by its transition (or next state) function, which can be interpreted as embodying instructions of a certain kind—e.g. "When receiving input 1 while in state s, print 0 and jump to state t". A difference between an animal capable of learning and a Turing machine is that the latter has a fixed transition function built into it, whereas the animal can build and rebuild by itself a whole set of transition functions. (More on the difference between brain and machines in Vol. 4, Ch. 5.)

Whereas some rules are arbitrary, others are justifiable. The former are conventions, such as the rule that judges should wear black robes, and women discriminated against. A rule can be justified in either of two ways: practically (by success) or theoretically (by its compatibility with some theory). Thus until recently the medical rule "If you have a headache take an aspirin" was justified only practically. The theoretical justification of a rule consists in showing, with the help of some theory, *why* it is efficient. We call such rules *grounded.* And if the grounding theory is (sufficiently) true, the corresponding rules will be said to be *well grounded*; they may also be called *nomopragmatic* (instead of nomological) statements.

Take for example the so-called phase rule of thermodynamics: $C+2=P+F$, where C is the number of components, P the number of phases, and F the number of degrees of freedom of a system in equilibrium. Notwithstanding its name it is a law, for it belongs to a theory and concerns physical or chemical systems, not our actions on them. However, like many another law, the phase "rule" can be used to obtain certain desired effects. Thus, by increasing (decreasing) C we may cause an increase (decrease) in P. In other words, this law statement is the ground for a number of rules proper. All the rules employed in science and technology are supposed to be well grounded, for an empirical rule may be sheer superstition or, if effective, it may have unknown undesirable side effects. On the other hand the rules used or recommended by occultists, quacks, pseudoscientists, and unscientific ideologists are not well grounded.

The rule for obtaining a nomopragmatic statement from a nomological one is simple: Substitute a pragmatic expression for at least one of the expressions occurring in the law formula. Table 9.1 exhibits a few examples.

The relation between well grounded rules and the corresponding laws is

TABLE 9.1
Obtaining well grounded rules from law statements.

Law Schema	*Rule Schema*
1. When A rises B falls.	1. To decrease B, increase A.
2. If A happens B occurs.	2. When A is done B is found to happen.
3. For every x and some y, Fxy.	3. For any given x, some y may be found such that Fxy.
4. The rate of change of F equals G.	4. To vary the rate of change of F regulate G.
5. Every system has assembled from its components under suitable circumstances.	5. To manufacture a system bring its components together under suitable circumstances.

as follows. Consider an elementary rule schema of the form "To attain G use M", or *G per M* for short, where G is a goal and M a means. We stipulate that this rule is *well grounded* if, and only if, there is a law according to which M brings about G. In other words, the rule is well grounded only in case there is a reasonably well-confirmed theory containing a statement of the form "If M then G", where "M" and "G" are interpreted, in the theory, as properties, states, or events.

Note that the very same law statement that provides a basis for *G per M* supports also the dual rule *Non-G per non-M*, i.e. "To prevent G from being the case abstain from using M". (The contraposition of the law statement is not the ground of the dual rule. And in the latter the expressions of the form 'not-X' do not mean "anti-X" but merely the absence of X.) For example, "The output of an economic system is an increasing function of both capital and labor" provides the ground for two rules: "To increase output increase capital or labor", and "To decrease output decrease capital or labor". This practical ambivalence of law statements cannot be accounted for by pragmatism, which mistakes laws for prescriptions.

To say that a rule is *based* or *grounded* on a law statement is not to say that it follows from the latter—if only because every law statement supports a pair of rules. This point is of interest not only to the philosophy of technology but also to all the normative disciplines. In particular, to demand that the ethical and legal norms be well grounded—e.g. on psychology and sociology—does not amount to requiring that they be deducible from them. In other words, even if we wish the normative disciplines to be based on science we cannot reduce them to the latter. (This point will be discussed in detail in Vol. 7.)

Laws and rules may combine in yet another way, namely as rules helping or blocking the formulation of hypotheses expected to attain the rank of law. One such blocking rule is the commandment that scientific formulas ought to involve only functions representing observable properties. (Unfortunately for this rule all the basic concepts of the fundamental physical theories, in particular those of field intensity, Lagrangian, and state function, represent unobservables.) Another such rule, but this time a fertile one, is the correspondence principle: "Every new theory should contain, in some limit or other, the true, or approximately true, formulas of its predecessors". A third is the general covariance principle: "Every basic law ought to be frame-independent, i.e. to remain invariant under changes in frame of reference". This rule instructs us to get rid of constants specifying particular (e.g. egocentric) facts, and it has been a powerful stimulus to the construction of coordinate-free geometries. A fourth is the rule of equipresence, widely used in contemporary thermomechanics: "A magnitude present as an independent variable in one constitutive relation should be so present in all". This rule suggests including in every proposed constitutive relation all the variables occurring in other formulas of the same area.

Neither of the above principles represents a law of nature: all of them are heuristic rules concerning scientific hypotheses and theories. We call them *metanomological* statements. Clearly, they are rules or norms, hence efficient, inefficient, or counterproductive but not true or false. They are rules to be justified exclusively by their success in suggesting or eliminating hypotheses aiming at capturing objective patterns: they cannot be put to the test in the laboratory or in the field. On the other hand the statement that the basic equations of classical mechanics remain invariant under a transformation of coordinates mirroring an exchange of inertial (Galilean) frames of reference, is a metanomological statement that can be proved with pencil and paper. The same holds for the statement of the Lorentz invariance of the Maxwell equations, and of the charge-parity-time reversal invariance of all the basic local field theories. These metatheoretical statements are true, and their truth can be checked by purely conceptual operations, so they do not represent any real patterns. So much so that one can build any number of theories abiding by such principles and having nothing to do with reality. (For more on metanomological statements see Bunge (1963, 1967a).)

This concludes our treatment of laws and rules, subjects that, despite their centrality, are neglected by most philosophers—or, worse, they are mixed up with extraneous topics such as induction, counterfactuals, and possible worlds.

4. CONCLUDING REMARKS

Most of us go through life without ever encountering well-constructed classifications or hypothetico-deductive systems. We get by with vague classifications that are neither exclusive nor exhaustive, and with stray conjectures or tangled bodies of propositions that we dignify with the label 'theory'. In Antiquity only a handful of thinkers, such as Aristotle, Euclid and Archimedes, proposed reasonable classifications or well-built theories. In the Middle Ages only a few theologians, in the first place Thomas Aquinas, constructed hypothetico-deductive theories—albeit speculative ones. And in the Modern period there was no general theory before Newton's, and no general and correct classification before Linnaeus'. But from then on classifications and theories flourished and multiplied. In particular, theory has become so central to science and technology, that any discipline that lacks theories is properly regarded as nonscientific or, at best, protoscientific.

However, the careful and systematic analysis of classifications and theories did not begin until the 19th century, in the first place by Whewell (1847). Before him philosophers had no adequate conception of theories. Thus Bacon cared for none and explicitly rejected the few scientific theories of his time, and the British empiricists ignored factual theories altogether. Descartes and Leibniz thought scientific theories were a priori mathematical structures, and never thought it necessary to subject them to experimental tests. Kant did conceive of scientific theory as a synthesis of reason and experience—but the wrong kind of synthesis. Indeed he thought theories can account for appearance not reality, and he failed to understand the scientific theories of his time for want of mathematical knowledge.

We are better placed than our ancestors to analyze the logical structure of classifications and theories, because in the meantime logicians and mathematicians have forged the necessary tools of analysis, such as the predicate calculus, the theories of sets, lattices, filters, and ideals, model theory, and the calculus of theories. (See Tarski, 1956; Rasiowa and Sikorski, 1963; Bell and Slomson, 1969.) Yet, all this concerns structure not content. Accordingly only logicians and mathematicians have attained an adequate (though surely not final) understanding of their own systematizations. Many scientists, technologists and philosophers are often still confused or wrong about the nature of scientific and technological systematizations. This is partly explainable by the intrinsic difficulty of the subject: not only the formal structure or formalism but also the content is to be accounted for. (Hilbert said that he found physics much more difficult than mathematics.) But there is also the weight of tradition: mathematics has always been helped by rationalism and never been seriously hampered by empiricism. On the other hand factual science, technology and philosophy are still being hindered by empiricism.

The misunderstanding and even mistrust of theory among many philosophers is apparent in the half-century old debate on the status of theoretical concepts. Many are still worried by the fact that theories happen to contain theoretical concepts. The less sophisticated among them get away by decreeing that there is no problem: that every scientific theory concerns observations or measurements, not autonomously existing things (e.g. Simon and Groen, 1973). But of course this operationist reform (or rather destruction) does not work, because (*a*) measurement instruments are designed and operated with the help of theories, (*b*) we often find it necessary to correct measurement results in the light of theories, and (*c*) the basic concepts of the most advanced theories happen not to represent observable things or properties. (See e.g. Bunge, 1973a.)

Other philosophers recognize the existence of theoretical concepts but try to eliminate them by various ingenious techniques, of which the following are the best known. (*a*) Define the theoretical concepts in terms of observational ones, the way Mach tried to define "mass" as a ratio of accelerations. (*b*) Construct Ramsey sentences with the help of the empirical concepts of the theory (e.g. Carnap, 1966). (*c*) Pretend that the theory is composed of formulas that are either observational or theoretical, and apply Craig's theorem (Craig, 1956). None of these techniques work. Mach's because it presupposes the very law statements used to "define" the suspect concepts. (See Bunge, 1966). Ramsey's because the theoretical concepts must have been introduced, either as primitives or as defined concepts, before the postulates of the theory can be stated and conjoined to form a Ramsey sentence. And Craig's technique fails because it rests on the false assumption that it is possible to partition the set of formulas of a theory into observational and theoretical, while actually no factual theories contain strictly observational formulas. And all three

TABLE 9.2
Characterization of epistemologies with respect to their theses on the representability of scientific systematizations.

Epistemology	*Thesis*
Conventionalism	Do not represent: just colligate data.
Empiricism	Represent human experiences.
Pragmatism	Do not represent: are just tools for action.
Naive realism	Represent directly and faithfully.
Critical realism	Represent indirectly and, at best, approximately.

theory demolition techniques fail because they ignore that factual theories happen to be about concrete things existing whether we know it or not, and having properties that can be represented only by sophisticated concepts. (For further criticisms see Bunge (1967a, 1973a, 1973b), and Tuomela (1973).) This view, critical realism, contrasts with the alternative views as shown in Table 9.2.

Finally let us define the reference, extension, scope, and power of a factual theory, and let us evaluate theorizing in general. The *reference* (class) of a theory is of course the collection of its referents. These are not only the actual and investigated ones but all of the possible things of a kind—e.g. all the possible biopopulations in the case of the genetics of populations. (See Vol. 1, Ch. 2.) On the other hand the *extension* of a theory T may be defined as the collection of referents of T for which T has been confirmed reasonably well, i.e. for which T holds with good approximation. Note that the extension of a theory is a subset of its reference, and that it is a variable collection rather than a fixed set, for the incoming empirical information is likely to expand or shrink, in the course of time, the range of things that are adequately covered by the theory.

The *scope* of a theory may be defined as the set of events it allows or regards as lawful, i.e. the event space it assigns to its referents. (The event space of things of kind K is properly included in the cartesian product of the state space S_K by itself—i.e. it is the set of pairs of states compatible with the laws of things of kind K. See Vol. 3, Ch. 5.) The more general a theory, the greater its scope. As a theory is adjoined subsidiary hypotheses accounting for details of its referents, its scope shrinks: it tells us more and more about less and less. We need theories of all scopes except the two extremes of complete restriction ("Nothing ever happens") and total permissiveness ("Anything goes"). The former extreme must be avoided because it leaves nothing to be investigated, the latter because it has no explanatory power and consecrates magic.

The *power* of a theory is measurable by the number of bits of knowledge it systematizes—in particular the number of precursor hypotheses it puts together—and the number of new items it introduces. Of all available scientific theories the quantum theory and the synthetic theory of evolution are the most powerful. But such power, though great, is not unlimited: no theory can embrace all the knowledge in a given field. The reason for this limitation is that a factual theory should contain no data. All we demand of theories is that they contain blanks that can be filled out with empirical data every time we apply them to solve particular problems. (Recall Section 2.2.)

As for the value of theories, different people value them differently. To theoretical scientists and rationalist philosophers, theory is at the very top, for it constitutes the best possible systematization, extrapolation, and exactification of a body of knowledge. To descriptive scientists, empiricists, philosophers, and laymen, theory is a matter of either indifference or embarrassment for overreaching experience, and so it is at the very bottom of their scale of values. And to workers in experimental science and technology, theory is right in the middle, for they are guided by theories and they know their discoveries or inventions motivate the building and correction of theories. The realist epistemologist values equally theory and experiment: to him it is a truism that theory without empirical support is sheer speculation, whereas empirical operation without theoretical backing is a waste of time.

This concludes the first part of our study of epistemology and methodology. The second part, titled *Understanding the World*, examines the ways we account for facts and check and evaluate our hypotheses, theories and proposals. It also discusses the varieties of human knowledge and of changes of it. Finally, it formulates and defends the general view resulting from our study, namely scientific realism.

APPENDIX 1

THE POWER OF MATHEMATICS IN THEORY CONSTRUCTION: A SIMPLE MODEL OF EVOLUTION

To illustrate the power of mathematics in the construction of theories consider the following simple-minded model of evolution representing the two main evolutionary "forces", namely random mutation and selection by the environment. (This model is quite different from the one in Vol. 4, Ch. 3, Section 3.4.) The evolving units may be populations of molecules, stars, organisms, or social systems: the theory does not specify the precise mutation or selection mechanisms. We assume that the components of such a population are subject to mutation, whether genic or of some other type, spontaneous or induced by environmental agents, but in any case "blind" or random, in the sense that the probability of a mutation is independent of its survival value for the component. We also suppose, for the sake of simplicity, that there is a single kind of mutation and that some of the mutants happen to be viable, so that in the course of time a second population emerges, namely that of mutants. Call the latter M and that of normals N. Suppose also that the mutations are irreversible and that the mutants themselves are stable, i.e. immune to any further mutations, so that they contribute neither to the normal population nor to any third group. Besides, assume that all the normals have the same probability p of originating mutating offspring, and that on the average each normal has a descendants, where $a > 1$. Finally, suppose that all of the viable individuals endure, regardless of their age, for a large number g of generations. Note that our model contains no less than six simplifying assumptions, most of which are at best only approximately true.

Our problems are (*a*) to trace the speciation process, i.e. the growth of the M population, and (*b*) to estimate the mutation parameter p from a count of the fraction of mutants in the total population after g generations. We start by noting that, since the probability of a mutation from one generation to the next is p, the probability that no such change occurs during the same period is $1-p$, and the probability that a normal individual will originate normal offspring at generation n will be

$$v_n = (1-p)^n.$$

Consequently the probability that an organism taken at random from the total population at generation n be a mutant, is the complement of v_n to unity, i.e.

$$\mu_m = 1 - (1 - p)^n.$$

Recall now that every normal individual produces, on the average, a descendants. Hence the average number of normals at generation n is

$$N_n = a^n v_n = a^n(1 - p)^n.$$

Therefore the total average of normals for all g generations is

$$N = \sum_0^g N_n = \sum_0^g a^n(1 - p)^n = \frac{(a(1 - p))^{g+1} - 1}{a(1 - p) - 1}.$$

To compute the corresponding number M of mutants produced at generation g assume that on the average every mutant produces b offspring, where b is in general different from a—in fact greater than a if the mutants are fitter than the normals. Hence the average number of mutants at generation n is

$$M_n = b^n \mu_n = b^n(1 - (1 - p)^n).$$

Consequently the average number of mutants at generation g is

$$M = \sum_0^g M_n = \sum_0^g b^n(1 - (1 - p)^n) = \sum_0^g b^n - \sum_0^g (b(1 - p))^n$$
$$= \frac{b^{g+1} - 1}{b - 1} - \frac{(b(1 - p))^{g+1} - 1}{b(1 - p) - 1}.$$

What can be observed directly are the totals N and M. From them we can estimate the mutation parameter p as follows. To simplify the calculations make the unrealistic pretense that the mutants are just as fit as the normals, i.e. set $a = b$, and use the previous results to form the fraction of normals:

$$\frac{N}{N + M} = \frac{\dfrac{(a(1 - p))^{g+1} - 1}{a(1 - p) - 1}}{\dfrac{a^{g+1} - 1}{a - 1}} \cong (1 - p)^{g+1},$$

since, for small p, $a(1 - p) \cong a$. Finally we obtain

$$p \cong 1 - \left(\frac{N}{N + M}\right)^{1/(g+1)}.$$

For example, for $p = 0.001$ and $g = 99$, the total fraction of mutants proves to be 0.1, an appreciable number. If the mutation probability is tenfold, the mutant population attains over one third of the total. But even a small mutation probability can be compensated for by a large fitness or fertility value. In other words, a favorable environment can make up for a small chance of an individual change.

Note the following points. First, the entire model involves only a couple of theorems in the elementary calculus of probability and a handful of algebraic manipulations. Second, every assumption is translated into mathematical terms, and every mathematical result is translated back into scientific terms, i.e. is assigned a factual interpretation. Third, the last two formulas, which relate the unobserved probability p to the observable frequency $N/(N+M)$, is characteristic of this theory. Different stochastic theories contain different relations between the two quantities—or none. (Therefore, it is hopeless to try and define probability as long term frequency, the way empiricists have tried. In general, theoretical concepts cannot be defined in terms of empirical ones or conversely. But they can be related.)

THE PROSE IDENTIFYING THE VARIABLES

Take the hypothesis that frustration builds up latent aggression (i.e. aggressiveness). This hypothesis may be taken as the precursor of a theory composed of mathematical formulas and semantic formulas. Let us proceed to build the simplest possible such theory.

We start by defining frustration as the percentage of unfulfilled desired acts, and postulate that latent aggression (or aggressive disposition) is proportional to manifest aggression. That is, we set

$$F=(D-P)/D \qquad \textit{Definition} \quad (1)$$

and

$$A=aN, \qquad \textit{Hypothesis} \quad (2)$$

where "D" and "P" stand for desired and performed acts of some kind respectively, while "N" stands for the total number of aggressions and "a" their intensity, assumed to be constant for the sake of simplicity.

We can now state the original (precursor) hypothesis in a definite fashion:

$$A=bF, \qquad \textit{Hypothesis} \quad (3)$$

where b is a positive real number. Obviously, this assumption of mere proportionality between aggressiveness and frustration is too simple to be true. But our purpose is semantic insight not factual truth.

Substituting (1) and (3) into (2) we obtain a testable consequence:

$$N=b(D-P)/aD. \qquad \textit{Theorem} \quad (4)$$

Finally we may collect the above formulas, restate them more carefully, and surround them with the requisite mathematical and semantical formulas. The result is this:

Primitive concepts
Set S of subjects
Number D of desired acts of a given kind per unit time
Number P of desired and performed acts of a given kind per unit time
Number N of aggressive acts per unit time

Defined concept

Frustration $F = (D - P)/D$

Postulates

Mathematical 1 $D, P: S \to \mathbb{N}$.
Mathematical 2 $A, a: S \to \mathbb{R}^+$
Mathematical 3 $b \in \mathbb{R}^+$
Factual 1 For every x in S, $F(x) = a(x)N(x)$.
Factual 2 For every x in S, $A(x) = bF(x)$.
Semantic 1 "$A(x)$" represents the aggressiveness of subject x.
Semantic 2 "$F(x)$" represents the frustration felt by subject x.
Semantic 3 "$a(x)$" represents the intensity of the aggressive acts committed by subject x.

Theorem

For every x in S, $N(x) = b[D(x) - P(x)]/a(x)D(x)$.

The same mathematical and factual axioms may occur in totally different contexts accompanied by corresponding different semantic axioms. And if the former are changed then the latter too may have to be modified. For example, instead of construing frustration and aggressiveness as numerical variables we may work with the sets of aggressive acts and desired acts, whether performed or frustrated. And we may then try a probabilistic version of the precursor hypothesis, such as, e.g.,

For any subject x, every a in A, every f in F, and every p in P,

$$Pr(a|f) = a(x)[1 - Pr(a|p)], \quad \text{with } a(x) > 1.$$

The semantic assumptions are now

$S1$ "$Pr(a|f)$" represents the propensity (tendency) of a subject to commit an aggressive act when he feels a frustrated desire f.
$S2$ "$Pr(a|p)$" represents the propensity (tendency) of a subject to commit an aggressive act when he has fulfilled desire p.
$S3$ "$a(x)$" represents the natural aggressiveness of the subject x.

The point is that every formalizations calls for its own set of semantic formulas, without which we would not know what it represents.

BIBLIOGRAPHY

Abelson, P. H. (1965). Relations of group activity to creativity in science. *Proc. Amer. Acad. Arts and Science* **94**: 603–614.

Ackoff, R. L. (1974). *Redesigning the Future: A Systems Approach to Societal Problems.* New York: Wiley-Interscience.

Ackoff, R. L. (1978). *The Art of Problem Solving.* New York: John Wiley and Sons.

Adams. J. L. (1980). *Conceptual Blockbusting*, 2nd ed. New York: W. W. Norton and Co.

Agassi, J. (1975). *Science in Flux.* Dordrecht and Boston: Reidel.

Alcock, J. (1981). *Parapsychology: Science or Magic?* Oxford: Pergamon Press.

Amari, S.-I., and A. Takeuchi (1978). Mathematical theory on formation of category detecting nerve cells. *Biol. Cybernetics* **29**: 127–136.

Aqvist, L. (1965). A New Approach to the Logical Theory of Interrogatives. Mimeo. University of Uppsala.

Asch, S. E. (1952). *Social Psychology.* New York: Prentice-Hall.

Aubrey, J. (1949). *Aubrey's Brief Lives.* O. Lawson Dick, Ed., London: Secker and Warburg.

Bandura, A. (1974). Behavior theory and the models of man. *Amer. Psychol.* **29**: 859–869.

Baranyi, A., and O. Fehér (1981). Synaptic facilitation requires paired activation of convergent pathways in the neocortex. *Nature* **290**: 413–415.

Barber, T. X. (1978). Hypnosis, suggestions, and psychosomatic phenomena: a new look from the standpoint of experimental studies. *Am. J. Clinical Hypnosis* **21**: 13–27.

Barber, T. X., and S. C. Wilson (1979). Guided imaging and hypnosis. In A. A. Sheikh and J. T. Shaffer, Eds., *The Potential of Fantasy and Imagination.* New York: Brandon House.

Barnes, B. (1982). *T. S. Kuhn and Social Science.* New York: Columbia University Press.

Bartlett, F. C. (1932). *Remembering.* Cambridge: Cambridge University Press.

Bartlett, F. (1958). *Thinking.* New York: Basic Books.

Battro, A. M., S. P. Netto, and R. J. A. Rozestraten (1976). Riemannian geometries of variable curvature in visual space. *Perception* **5**: 9–23.

Battro, A. M. (1979). Psicología, geometría y filosofía del espacio visual. *Revista Latinoamericana de Filosofía* v: 19–31.

Bechtereva, N. P. (1978). *The Neurophysiological Aspects of Human Mental Activity.* New York: Oxford University Press.

Becker, G. (1964). *The Human Capital*, 2nd ed. New York: Columbia University Press.

Beer, C. G. (1973). A view of birds. In A. Pick, Ed., *Minnesota Symposia on Child Psychology*, Vol. 7, pp. 47–86. Minneapolis: University of Minnesota Press.

Békesy, G. von (1967). *Sensory Inhibition.* Princeton: Princeton University Press.

Békesy, G. von (1968). Problems relating psychological and electrophysiological observations in sensory perception. *Perspectives in Biology and Medicine* **11**: 179–194.

Bell, J. L., and A. B. Slomson (1969). *Models and Ultraproducts.* Amsterdam: North-Holland.

Bellugi, U., and E. S. Klima (1979). Language: Perspectives from another modality. In *Brain and Mind*, Ciba Foundation Symposium 69, pp. 99–117.

Berger, P. L., and T. Luckmann (1966). *The Social Construction of Reality*. New York: Doubleday.

Berlyne, D. E. (1954). A Theory of Human Curiosity. *Brit, J. Psychol.* **45**: 180–191.

Bertalanffy, L. von (1955). An essay on the evolution of categories. *Phil. Science* **22**: 243–263.

Bindra, D. (1974). A motivational view of learning, performance, and behavior modification. *Psychol. Rev.* **81**: 199–213.

Bindra, D. (1976). *A Theory of Intelligent Behavior*. New York: Wiley Interscience.

Bindra, D., Ed. (1980). *The Brain's Mind: A Neuroscience Perspective on the Mind–Body Problem*. New York: Gardner Press.

Bindra, D. (1981). Ape language. *Science* **211**: 86.

Bindra, D. (1982). Cognitivism: Its origin and future in psychology. *Annals of Theoretical Psychology*, forthcoming.

Birch, H. G. (1945). The relation of previous experience to insightful problem-solving. *J. Comparative Psychology* **38**: 367–383.

Birch, H. G., and H. S. Rabinowitz, (1951). The negative effect of previous experience on productive thinking. *J. Experimental Psychology* **41**: 121–125.

Black, M. (1962). *Models and Metaphors*. Ithaca, New York: Cornell University Press.

Bliss, T. V. P. (1979). Synaptic plasticity in the hippocampus. *Trends in Neurosciences* **2**: 42–45.

Bloor, D. (1977). The regulatory function of language. In Morton and Marshall, Eds. (1977), pp. 73–98.

Bohr, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **48**: 686–702.

Boltzmann, L. (1905). *Populäre Schriften*. Leipzig: Barth.

Bolzano, B. (1837). *Wissenschaftslehre*, 4 vols. Repr.: Lepzig, Meiner, 1929.

Botha, R. P., and W. K. Winkler (1973). *The Justification of Linguistic Hypotheses*. The Hague: Mouton.

Botkin, J. W., M. Elmandjra, and M. Malitza (1979). *No Limits to Learning*. Oxford: Pergamon Press.

Bouchard, T. J., Jr., and M. McGue (1981). Familial studies of intelligence: a review: *Science* **212**: 1055–1059.

Boudon, R. (1967). *L'analyse mathématique des faits sociaux*. Paris: Librairie Plon.

Brewer, D. (1980). On the theory and practice of innovation. *Technology in Society* **2**: 337–363.

Bridgman, P. W. (1927). *The Logic of Modern Physics*. New York: Macmillan.

Brines, M. L., and J. L. Gould (1979). Bees have rules. *Science* **206**: 571–573.

Brown, H. I. (1977). *Perception, Theory and Commitment. The New Philosophy of Science*. Chicago: Precedent Publ. Inc.

Bruner, J. S., J. J. Goodnow, and G. Austin (1959). *A Study of thinking*. New York: John Wiley and Sons.

Bunge, M. (1959a). *Causality*. Cambridge, Mass.: Harvard University Press. Rev. ed.: *Causality in Modern Science*. New York: Dover, 1979.

Bunge, M. (1959b). *Metascientific Queries*. Evanston, Ill.: Charles C. Thomas.

Bunge, M. (1962a). *Intuition and Science*. Englewood Cliffs, N.J.: Prentice-Hall. Repr.: Westport, Conn.; Greenwood Press, 1975.

Bunge, M. (1964). Phenomenological theories. In M. Bunge, Ed., *The Critical Approach*, pp. 234–254. New York: The Free Press.

Bunge, M. (1966). Mach's critique of Newtonian mechanics. *American Journal of Physics* **34**: 585.

Bunge, M. (1967a). *Scientific Research I. The Search for System.* New York: Springer-Verlag.

Bunge, M. (1967b). *Scientific Research II. The Search for Truth.* New York: Springer-Verlag.

Bunge, M. (1967c). *Foundations of Physics.* New York: Springer-Verlag.

Bunge, M. (1967d). The structure and content of a physical theory. In M. Bunge, Ed., *Delaware Seminar in the Foundations of Physics*, pp. 15–27. New York: Springer-Verlag.

Bunge, M. (1968). Scientific laws and rules. In R. Klibansky, Ed., *Contemporary Philosophy* II, pp. 128–140. Firenze: La Nuova Italia Editrice.

Bunge, M. (1970). Virtual processes and virtual particles: real or fictitious? *Intern. J. Theoretical Physics* **3**: 507–508.

Bunge, M. (1971). A mathematical theory of dimensions and units. In M. Bunge, Ed., *Problems in the Foundations of Physics*, pp. 1–16. New York: Springer-Verlag.

Bunge, M. (1973a). *Philosophy of Physics.* Dordrecht: Reidel.

Bunge, M. (1973b). *Method, Model and Matter.* Dordrecht: Reidel.

Bunge, M., Ed. (1973d). *Exact Philosophy.* Dordrecht: Reidel.

Bunge, M. (1973e). Review of Suppes' (1970). *Brit. J. Phil. Sci.* **24**: 409–410.

Bunge, M. (1974a). *Sense and Reference* (Treatise, Vol. 1). Dordrecht: Reidel.

Bunge, M. (1974b). *Interpretation and Truth* (Treatise, Vol. 2). Dordrecht: Reidel.

Bunge, M. (1977a). *The Furniture of the World* (Treatise, Vol. 3). Dordrecht and Boston: Reidel.

Bunge, M. (1977b). Levels and reduction. *Am. J. Physiol.* **233**: R75–82.

Bunge, M. (1977d). The philosophical richness of technology. In F. Suppe and P. D. Asquith, Eds., *PSA 1976*, Vol. 2, pp. 153–172. East Lansing, Mich.: Philosophy of Science Assn.

Bunge, M. (1977f). General systems and holism. *General Systems* **XXII**: 87–90.

Bunge, M. (1978b). A model of evolution. *App. Math. Modell.* **2**: 201–204.

Bunge, M. (1980a). *The Mind-Body Problem.* Oxford and New York: Pergamon Press.

Bunge, M. (1980b). *Ciencia y desarrollo.* Buenos Aires: Siglo Veinte.

Bunge, M. (1981a). *Scientific Materialism.* Dordrecht and Boston: Reidel.

Bunge, M. (1981b). Analogy between systems. *Intern. J. General Systems* **7**: 221–223.

Bunge, M. (1981c). Four concepts of probability. *Appl. Math. Modell.* **5**: 306–312.

Bunge, M. (1981e). Review of Fleck (1935). *Behav. Sci.* **26**: 178–180.

Bunge, M. (1982a). The revival of causality. In G. Fløistad, Ed., *Contemporary Philosophy*, Vol. 2, pp. 135–155. The Hague: Martinus Nijhoff.

Bunge, M. (1982b). *Economía y filosofía.* Madrid: Tecnos.

Bunge, M. (1983) *Lingüística y filosofía.* Barcelona: Ariel.

Calvin, W. H. and G. A. Ojemann (1980). *Inside the Brain.* New York: New American Library.

Campbell, D. T. (1959). Methodological suggestions from a comparative psychology of knowledge processes. *Inquiry* **2**: 152–182.

Campbell, D. T. (1974). Evolutionary epistemology. In Schilpp, Ed., pp. 413–463.

Campbell, D. T. (1977). *The William James Lectures*, Harvard University. Preprint courtesy of Professor Campbell.

Campbell, D. T. (1979). A tribal model of the social system vehicle carrying scientific knowledge. *Knowledge* **1**: 181–217.

Čapek. M. (1968). Ernst Mach's biological theory of knowledge. *Synthese* **18**: 171–191.

Carnap, R. (1966). *Philosophical Foundations of Physics.* M. Gardner. Ed. New York: Basic Books.

Carterette, E. C., and M. P. Friedman, Eds. (1976). *Handbook of Perception*, Vol. VII. New York: Academic Press.

Causey, R. L. (1977). *Unity of Science*. Dordrecht-Boston: Reidel.

Chapanis, N. P., and A. Chapanis (1954). Cognitive dissonance: five years later. *Psychol Bull.* **61**: 1–22.

Chapman, L. J., and J. P. Chapman (1959). Atmosphere effect re-examined. *J. Exp. Psychol.* **58**: 220–226.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon Books.

Chomsky, N. (1980). *Rules and Representations*. New York: Columbia University Press.

Claparède, E. (1934). La genèse de l'hypothèse. Étude experimentale. *Archives de psychologie* **24**: 1–155.

Clark, E. V. (1977). First language acquisition. In Morton and Marshall, Eds., pp. 1–72.

Cohen, N. L., and L. R. Squire (1980). Preserved learning and retention of pattern-analyzing skill in amnesia: dissociation of knowing how and knowing that. *Science* **210**: 207–209.

Cole, S., J. R. Cole, and G. A. S. Simon (1981). Chance and consensus in peer review, *Science* **214**: 881–886.

Coleman, J. S. (1980). The structure of society and the nature of social research. *Knowledge* **1**: 333–350.

Conel, J. L. (1939–1967). *The Postnatal Development of the Human Cerebral Cortex*, 8 volumes. Cambridge, Mass.: Harvard University Press.

Craig, W. (1956). Replacement of auxiliary expressions. *Philosophical Review* **65**: 38–55.

Currie, G. (1978). Popper's evolutionary epistemology. *Synthese* **37**: 413–431.

Davenport, L. D. (1979). Superstitious bar pressing in hippocampal and septal rats. *Science* **205**: 721–723.

Davidson, D. (1963). Actions, reasons, and causes. *J. Phil.* **60**: 685–700.

Dawson, M. E., and J. J. Furedy (1976). The role of awareness in human differential autonomic classical conditioning: the necessary-gate hypothesis. *Psychophysiology* **13**: 50–53.

Denenberg, V. (1979). Paradigms and paradoxes in the study of behavioral development. In E. Thoman, Ed., *Origins of the Infant's Social Responsiveness*, pp. 251–289. Hillsdale, N.J.: L. Erlbaum Assoc.

Denison, E. F. (1962). *The Sources of Economic Growth in the United States*. New York: Committee for Economic Development.

Dickinson, A. (1980). *Contemporary Animal Learning Theory*. Cambridge: Cambridge University Press.

Dickinson, A., and N. J. Mackintosh (1978). Classical conditioning in animals. *Annual Rev. of Psychology* **29**: 587–612.

Dodwell, P. C. (1975). Contemporary theoretical problems in seeing. *Handbook of Perception*, Vol. V: 57–77. New York: Academic Press.

Dreistadt, R. (1969). The use of analogies and incubation in obtaining insights in creative problem solving. *Journal of Psychology* **71**: 159–175.

Duhem, P. (1914). *Essai sur la notion de théorie physique*, 2nd ed. Paris: Hermann.

Duncker, K. (1945). On problem-solving. *Psychological Monograph* 58, No. 270.

Dyson, F. J. (1958). Innovation in physics. *Scientific American* **199**: 3.

Eberle, R. D. Kaplan, and R. Montague (1961). Hempel and Oppenheim on explanation. *Phil. Sci.* **28**: 418–428.

Eccles, J. C. (1978). *The Human Mystery*. New York: Springer-Verlag.

Eccles, J. C. (1980). *The Human Psyche*. Berlin, Heidelberg, New York: Springer-Verlag.

Einhorn, H. J., and R. M. Hogarth (1978). Confidence in judgment: persistence in the illusion of validity. *Psychol. Rev.* **85**: 395–416.

Ellis, A. E. (1963). An introduction to the principles of scientific psychoanalysis. In S. Rachman, Ed., *Critical Essays on Psychoanalysis*: Oxford: Pergamon Press.

Evans, E. F., and I. C. Whitfield (1964). Classification of unit responses in the auditory cortex of the unanesthesized and unrestrained cat. *J. Physiol.* **171**: 476–493.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. I, 3rd. ed. New York: Wiley.

Ferrater Mora, J. (1967). *El ser y el sentido*. Madrid: Revista de Occidente.

Festinger, L. (1957). *Cognitive Dissonance*. Stanford: Stanford University Press.

Feyerabend, P. K. (1975). *Against Method*. Repr., London: Verso, 1978.

Feyerabend, P. K. (1981). *Philosophical Papers*, 2 Vols., Cambridge: Cambridge University Press.

Finetti, B. de (1972). *Probability, Induction and Statistics*. New York: John Wiley.

Fleck, L. (1935). *Genesis and development of a Scientific Fact*. Transl. F. Bradley and T. J. Trenn. Chicago: University of Chicago Press, 1979.

Flohr, H., and W. Precht, Eds. (1981). *Lesion-Induced Neuronal Plasticity in Sensorimotor Systems*. New York: Springer-Verlag.

Fodor, J. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

Fodor, J. (1981). The mind–body problem. *Scientific American* **244, 1**: 114–123.

Garcia, J. (1981). Tilting at the paper mills of Academe. *American Psychologist* **36**: 149–158.

Gardner, R. A., and B. T. Gardner (1969). Teaching sign language to a chimpanzee. *Science* **165**: 664–672.

Ghiselin, M. T. (1981). Categories, Life, and thinking. *Behav. and Brain Sci.* **4**: 269–283.

Gibson, J. (1966). *The Senses Considered as Perceptual Systems*. Boston: Houghton Mifflin.

Gibson, J. (1979). *The Ecological Approach to Visual Perception*. Boston: Houghton Mifflin.

Gilbert, C. D., and T. Wiesel (1979). Morphology and intracortical projections of functionally characterised neurones in the cat visual cortex. *Nature* **280**: 120–125.

Glass, L. (1975). Classification of biological networks by their qualitative dynamics. *J. Theoretical Biology* **54**: 85–107.

Goddard, G. V. (1980). Component properties of the memory machine: Hebb revisited. In Jusczyk and Klein, Eds., pp. 231–247.

Goody, J. (1977). *The Domestication of the Savage Mind*. Cambridge: Cambridge University Press.

Gould, S. J. (1978). Morton's ranking of races of cranial capacity. *Science* **200**: 503–509.

Greenberg, J. H., M. Reivich, A. Alavi, P. Hand, A. Rosenquist, W. Rintelmann, A. Stein, R. Tusa, R. Dann, D. Christman, J. Fowler, B. McGregor, and A. Wolf (1981). Metabolic mapping of functional activity in human subjects, etc. *Science* **212**: 678–680.

Greenwell, J. R. (1980). Academia and the occult: An experience at Arizona. *The Skeptical Inquirer* **5**: 39–45.

Gregory, R. (1970). *The Intelligent Eye*. New York: MacGraw-Hill.

Gruber, H. E. (1981). On the relation between "aha experiences" and the construction of ideas. *History of Science* **19**: 41–59.

Hájek, P., and T. Hávranek (1978). *Mechanizing Hypothesis Formation*. New York: Springer-Verlag.

Hansel, C. E. M. (1980). *ESP and Parapsychology*. Buffalo, New York: Prometheus Books.

Hanson, N. R. (1958). *Patterns of Discovery*. Cambridge: Cambridge University Press.

Harnad, S. R., H. D. Steklis, and J. Lancaster, Eds. (1976). *Origins and Evolution of Language and Speech. Annals of the New York Academy of Sciences*, Vol. 280.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Hebb, D. O. (1951). The role of neurological ideas in psychology. *Journal of Personality* **20**: 39–55.

Hebb, D. O. (1966). *A Textbook of Psychology*. Philadelphia: W. B. Saunders.

Hebb, D. O., W. E. Lambert, and G. R. Tucker (1971). Language, thought and experience. *Modern Language Journal* **55**: 212–222.

Hebb, D. O. (1976). Physiological learning theory. *J. Abnormal Child Psychol.* **4**: 309–314.

Hebb, D. O. (1980). *Essay on Mind*. New York: L. Erlbaum Assoc.

Hécaen, H., and M. L. Albert (1978). *Human Neuropsychology*. New York: Wiley and Sons.

Held, R., and A. Hein (1963). Movement-produced stimulation in the development of visually guided behavior. *J. Compar Physiol. Psychol.* **56**: 872–876.

Hellman, R. H. G., Ed. (1980). *Nonlinear Dynamics. Annals New York Acad. Sci.*, v. 357.

Helmholtz, H. von (1873). *Popular Lectures on Scientific Subjects*. London: Longmans, Green and Co.

Helmholtz, H. von (1879). *Die Tatsachen in der Wahrnehmung*. Berlin: Hirschwald.

Hempel, C. G. (1962). Deductive-nomological *vs.* statistical explanation. In H. Feigl and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* III, pp. 98–169.

Hennessy, J. W., and S. Levine (1979). Stress, arousal, and the pituitary–adrenal system: a psychoendocrine hypothesis. *Progress in Psychobiology and Physiological Psychology* **8**: 133–178.

Hesse, M. (1966). *Models and Analogies in Science*. Notre Dame, Ind.: University of Notre Dame Press.

Hessen, B. (1931). The social and economic roots of Newton's "Principia". In *Science at the Cross Roads* pp. 149–212. London: Frank Cass, 1971.

Hilbert, D. (1918). Axiomatisches Denken, *Mathematische Annalen* **78**: 405–415.

Hinde, R. A. (1970). *Animal behavior. A Synthesis of Ethology and Comparative Physiology*, 2nd ed. New York: McGraw-Hill.

Hintikka, J. (1962). *Knowledge and Belief*. Ithaca, New York: Cornell University Press.

Hintikka, J. (1969). *Models for Modalities*. Dordrecht: Reidel.

Hoffman, W. C. (1966). The Lie Algebra of Visual Perception. *J. Math. Psychol.* **3**: 65–98.

Holton, G. (1973). *Thematic Origins of Scientific Thought*. Cambridge. Mass.: Harvard University Press.

Holton, G. (1978). *The Scientific Imagination. Case Studies*. Cambridge. Mass.: Harvard University Press.

Hubel, D. H., and T. N. Wiesel (1962). Receptive fields, binocular interaction, and functional architecture in the cat's visual cortex *J Physiol.* **160**: 106–154.

Hubel, D. H., and T. H. Wiesel (1977). Functional architecture of macaque visual cortex. *Proc. Royal Soc. (London) B*, **198**: 1–59.

Hume, D. (1739). *A Treatise of Human Nature*. L. A. Selbidge, Ed. Oxford: Clarendon Press.

Hunter, W. S. (1929). The sensory control of the maze habit in the white rat. *J. Genetic Psychol* **36**: 505–537.

Inhelder, B., and J. Piaget (1958). *The Growth of Logical Thinking from Childhood to Adolescence*. New York: Basic Books.

James, W. (1890). *Principles of Psychology*, 2 volumes. Repr. New York: Dover, 1950.

James, W. (1907). *Pragmatism*. New York: Meridian, 1955.

Jantsch, E. (1972). *Technological Planning and Social Futures.* London: Cassell.

Jerison, H. J. (1973). *Evolution of the Brain and Intelligence.* New York: Academic Press.

Johnson-Laird, P. N., and P. C. Wason, Eds. (1977). *Thinking: Readings in Cognitive Science.* Cambridge: Cambridge University Press.

Jusczyk, P. W., and R. M. Klein, Eds. (1980). *The Nature of Thought. Essays in Honor of D. O. Hebb.* Hillsdale, N. J.: Lawrence Erlbaum.

Kandel, E. R., and J. H. Schwartz, Eds. (1981). *Principles of Neural Science.* New York: Elsevier/North Holland.

Kant, I. (1787). *Kritik der reinen Vernunft,* 2nd ed. R. Schmidt, Ed., Hamburg: Meiner, 1930.

Kapitza, P. L. (1979). Plasma and the controlled thermonuclear reaction. *Science* **205**: 959–964.

Karmiloff-Smith, A., and B. Inhelder (1974/5). "If you want to get ahead, get a theory". *Cognition* **3**: 195–212.

Katona, G. (1939). *Organizing and Memorizing.* New York: Columbia University Press.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money.* Collected Writings Vol. VII. London: Macmillan and Cambridge University Press 1973.

Kinsbourne, M. (1971). The minor cerebral hemisphere as a source of aphasic speech. *Arch. Neurol.* **25**: 302–306.

Krechevsky, I. (1932). "Hypotheses" versus "chance" in the pre-solution period in sensory discrimination learning. University of California Publications in Psychology, Vol. 6, No. 3.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kuhn, T. S. (1974). Second thoughts on paradigms. In F. Suppe, Ed., *The Structure of Scientific Theories,* pp. 459–482. Urbana, Ill.: University of Illinois Press.

Kyburg, H. J., Jr. (1961). *Probability and Logic of Rational Belief.* Middleton, Conn.: Wesleyan University Press.

Lakatos, I. (1970). Falsification and the methodology of scientific research programmes. In Lakatos and Musgrave, Eds., pp. 91–195.

Lakatos, I. (1978). *Philosophical Papers,* 2 vols. J. Worrall and G. Currie, Eds., Cambridge: Cambridge University Press.

Lakatos, I., and A. Musgrave, Eds. (1970). *Criticism and the Growth of Knowledge.* Cambridge: University Press.

Lalande, André (1938). *Vocabulaire technique et critique de la philosophie,* 3 vols., 4th ed. Paris: Alcan.

Lambert, W. E. (1981). Bilingualism and language acquisition. *Annals of the New York Academy of Science* **379**: 9–22.

Lashley, K. (1929). *Brain Mechanisms and Intelligence.* Chicago: University of Chicago Press.

Latour, B., and S. Woolgar (1979). *Laboratory Life. The Social Construction of Scientific Facts.* Beverly Hills, Cal.: Sage Publications.

Lederberg, J., and E. A. Feigenbaum (1968). Mechanization of inductive inference in organic chemistry. In B. Kleinmutz, Ed. *Formal Representation of Human Judgment,* pp. 187–218. New York: John Wiley and Sons.

Lehrer, K. (1974). *Knowledge.* Oxford: Clarendon Press.

Leibniz, G. W. von (1703). *Nouveaux essais sur l'entendement humain.* Paris: Flammarion, s.d.

Lenin, V. I. (1908). *Materialism and Empirio-Criticism.* London: Lawrence and Wishart, 1950.

Lenzen, W. (1980). *Glauben, Wissen und Wahrscheinlichkeit.* Wien–New York: Springer-Verlag.

Leonardo da Vinci (1492 or 1515). Moral Precepts for the Student of Painting. In J. P. Richter, Ed., *The Notebooks of Leonardo da Vinci.* J. P. Richter, Ed., 2 vols. London: Sampson Low, Marston, Searle and Rivington, 1883. Repr. New York: Dover, 1970.

Leuba, C. (1955). Toward some integration of learning theories: the concept of optimal stimulation. *Psychol. Reports* **1**: 27–33.

Levine, M. (1966). Hypothesis behavior by humans during discrimination learning. *J. Exper. Psychology* **71**: 331–338.

Levine, M. (1974). *A Cognitive Theory of Learning.* Hillside, N. J.: L. Erlbaum Assocs.

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation.* La Salle, Ill.: Open Court.

Lewis, W. (1955). *The Theory of Economic growth.* London: Allen and Unwin.

Lieberman, P. (1975). *On the Origins of Language: An Introduction to the Evolution of Human Speech.* New York: Macmillan.

Lindauer, M. (1971). *Communication Among Social Bees.* Cambridge, Mass.: Harvard University Press.

Lloyd, G. E. R. (1966). *Polarity and Analogy.* Cambridge: Cambridge University Press.

Locke, J. (1689). *An Essay Concerning Human Understanding.* London: Routledge and Kegan Paul, s.d.

Loehlin, J. C., and R. C. Nichols (1976). *Heredity, Environment, and Personality.* Austin: University of Texas Press.

Lovejoy, A. O. (1929). *The Revolt Against Dualism.* La Salle, Ill.: Open Court.

Luneburg, R. K. (1947). *Mathematical Analysis of Binocular Vision.* Princeton: Princeton University Press.

Luria, A. R. (1961). *The Role of Speech in the Regulation of Normal and Abnormal Behavior.* Oxford: Pergamon Press.

Luria, A. R. (1976). *Cognitive Development. Its Cultural and Social Foundations.* Cambridge, Mass.: Harvard University Press.

Mach, E. (1905). *Erkenntnis und Irrtum.* Leipzig: Barth.

Mach, E. (1910). *Populär-wissenschaftliche Vorlesungen*, 4th ed. Leipzig: Barth.

Mackey, M. C., and L. Glass (1977). Oscillation and chaos in physiological control systems. *Science* **197**: 287–289.

Masterton, R. B., C. B. G. Campbell, M. E. Bitterman, and N. Hotton, Eds. (1976a). *Evolution of Brain and Behavior in Vertebrates.* Hillsdale, N. J.: L. Erlbaum Assoc.

Masterton, R. B., W. Hodos, and H. J. Jerison, Eds. (1976b). *Evolution, Brain, and Behavior.* Hillsdale, N. J.: L. Erlbaum Assoc.

Maudsley, H. (1876). *The Physiology of Mind.* London: Macmillan.

Mayer, R. E. (1977). *Thinking and Problem-Solving.* Glenview, Ill.: Scott, Foresman and Co.

McCloskey, M., A. Caramazza, and B. Green (1980). Curvilinear motion in the absence of forces: naive beliefs about the motion of objects. *Science* **210**: 1139–1141.

McKenna, M. C. (1976). Comments on Radinsky's "Later mammal radiations". In R. B. Masterton *et al.*, Eds. *Evolution of Brain and Behavior in Vertebrates.* Hillsdale, N. J.: Lawrence Erlbaum Associates.

Medawar, P. B. (1969). *Induction and Intuition in Scientific Thought.* Philadelphia: American Philosophical Society.

Meier, N. R. F. (1931). Reasoning in humans. II. *J. Comparative Psychology* **12**: 181–194.

Mermin, N. D. (1981). Quantum mysteries for anyone. *J. Phil.* **78**: 397–408.

Merton, K. (1973). *The Sociology of Knowledge.* Chicago: University of Chicago Press.

Mill, J. S. (1875). *A System of Logic*, 8th ed. London: Longmans, Green, 1952.

Miller, G. A. (1967). *The Psychology of Communication.* New York: Basic Books.

Milner, P. (1970). *Physiological Psychology.* New York: Holt, Rinehart and Winston.

Minsky, M. (1977). Frame-system theory. In Johnson-Laird and Wason, Eds. pp. 355–376.

Modigliani, F. (1977). The monetarist controversy or, should we forsake stabilization policies? *Amer. Econ. Rev.* **67**: 1–19.

Morton, J., and J. C. Marshall, Eds. (1977). *Psycholinguistics Series 1: Developmental and Pathological.* London: Elk Science.

Moruzzi, G. and Magoun, H. W. (1949). Brain stem reticular formation and activation of the EEG. *Electroencephalography and Clinical Neurophysiology* **1**: 455–473.

Myrdal, G. (1969). *Objectivity in Social Research.* New York: Pantheon Books.

Myrdal, G. (1973). *Against the Stream.* New York: Pantheon Books.

Nagel, E. (1961). *The Structure of Science.* New York: Harcourt, Brace and World.

Neisser, U. (1963). The multiplicity of thought. *Brit. J. Psychol.* **54**: 1–14.

Neisser, U. (1976). *Cognition and Reality.* San Francisco: Freeman.

Neisser, U. (1980). *The Limits of Cognition.* In Jusczyk and Klein, Eds., pp. 115–132.

Nelson, L. (1912). The impossibility of the "theory of knowledge". In *Socratic Method and Critical Philosophy: Selected Essays.* Transl. T. K. Brown III. New Haven, Conn.: Yale University Press. Repr.: New York, Dover, 1965.

Newell, A., and H. Simon (1972). *Human Problem Solving.* Englewood Cliffs, N. J.: Prentice-Hall.

Nisbett, R., and L. Ross (1980). *Human Inference: Strategies and Shortcomings of Social Judgment.* Englewood Cliffs, N. J.: Prentice-Hall.

Noton, D., and L. Stark (1971). Scanpaths in the eye movements during pattern perception. *Science* **171**: 308–311.

Nottebohm, F. (1981). A brain for all seasons: cyclical anatomical changes in song control nuclei of the canary brain. *Science* **214**: 1368–1370.

Oakeshott, M. (1962). *Rationalism in Politics and other Essays.* London: Methuen.

O'Keefe, J., and L. Nadel (1978). *The Hippocampus as a Cognitive Map.* Oxford. Clarendon Press.

Olds, J. (1975). Mapping the mind onto the brain. In F. G. Worden, J. P. Swazey, and G. Adelman, Eds. (1975). *The Neurosciences: Paths of Discovery.* Cambridge, Mass.: MIT Press.

Patterson, F. (1978). Linguistic capabilities of a low-land gorilla. In F. C. C. Peng, Ed., *Sign Language Acquisition in Man and Ape.* Boulder Colo.: Westview Press.

Patterson, F. (1981). Ape language. *Science* **211**: 86–87.

Pavlov, I. P. (1927). *Conditioned Reflexes.* New York: Dover, 1960.

Pavlov, I. P. (1955). *Selected Works.* Moscow: Foreign Languages Publ. House.

Penfield, W., and T. Rasmussen (1950). *The Cerebral Cortex of Man.* New York: Macmillan.

Phelps, M., D. E. Kuhl, and J. C. Mazziotta (1981). Metabolic mapping of the brain's response to visual stimulation: studies in humans. *Science* **211**: 1445–1448.

Piaget, J. (1952). *The Origins of Intelligence in Children.* New York: International Universities Press.

Piaget, J. (1955). *The Language and Thought of the Child.* New York: Meridian.

Piaget, J. (1967). *Biologie et connaissance.* Paris: Gallimard.

Piaget, J. (1972). Intellectual evolution from adolescence to adulthood. *Human Development* **15**: 1–12.

Piaget, J. (1976a). *The Psychology of Intelligence.* Totowa, N. J.: Adams.

Piaget, J. (1976b). *Le comportement, moteur de l'évolution.* Paris: Gallimard.

Piaget, J., and B. Inhelder (1967). *The Child's Conception of Space.* New York: Norton.

Pincus, J. H., and G. J. Tucker (1978). *Behavioral Neurology,* 2nd ed. New York: Oxford University Press.

Polanyi, M. (1958). *Personal Knowledge.* Chicago, University of Chicago Press.

Pólya, G. (1954). *Mathematics and Plausible Reasoning,* 2 vols. Princeton: Princeton University Press.

Pólya, G. (1957). *How to Solve it*? New York: Doubleday Anchor Books.

Popper, K. R. (1957). *The Poverty of Historicism.* London: Routledge and Kegan Paul.

Popper, K. R. (1959). *The Logic of Scientific Discovery.* London: Hutchinson.

Popper, K. R. (1963). *Conjectures and Refutations.* New York: Basic Books.

Popper, K. R. (1972). *Objective Knowledge.* Oxford: Clarendon Press.

Popper, K. R. (1974). Intellectual Autobiography. In P. Schilpp, Ed., *The Philosophy of Karl R. Popper.*

Popper, K. R., and J. C. Eccles (1977). *The Self and its Brain.* New York: Springer-Verlag.

Premack, D. (1971). Language in chimpanzee? *Science* **172**: 808–822.

Putnam, H. (1960). Minds and machines. In S. Hook, Ed., *Dimensions of Mind,* pp. 148–179. New York: New York University Press.

Putnam, H. (1975). *Philosophical Papers,* 2 vols. Cambridge: Cambridge University Press.

Pylyshyn, Z. W. (1980). Computation and cognition: issues in the foundations of cognitive science, *The Behavioral and Brain Sciences* **1**: 111–132.

Ramsey, F. P. (1931). *The Foundations of Mathematics.* London: Routledge and Kegan Paul.

Randi, J. (1979). A controlled test of dowsing abilities. *The Skeptical Inquirer* **4**: 16–29.

Rasiowa, H., and R. Sikorski (1963). *The Mathematics of Metamathematics.* Warsaw: Panstwowe Wydawnictwo Naukowe.

Ratliff, F. (1965). *Mach Bands: Quantitative Studies on Neural Networks of the Retina.* S. Francisco: Holden-Day.

Ratliff, F. (1971). Illusions in man and his instruments. *J. Phil.* **LXVIII**: 591–597.

Rescher, N. (1976). *Plausible Reasoning.* Assen-Amsterdam: Van Gorcum.

Rescher, N. (1977). *Methodological Pragmatism.* New York: New York University Press.

Restle, F. (1976). The selection of strategies in cue learning. *Psychol. Rev.* **69**: 329–343.

Rignano, E. (1923). *The Psychology of Reasoning.* Transl. W. A. Holl London: Kegan Paul, Trench, Trubner and Co.

Roche, M. (1976). Factors governing the scientific and technological development of a country. *Scientia* **111**: 75–84.

Rorty, R. (1979). *Philosophy and the Mirror of Nature.* Princeton: Princeton University Press.

Rosch, E. (1974). Linguistic relativity. In A. Silverstein, Ed., *Human Communication: theoretical Perspectives.* Hillsdale, N. J.: L. Erlbaum Assoc.

Russell, B. (1914). *Our Knowledge of the External World.* London: George Allen and Unwin.

Russell, B. (1918). *Mysticism and Logic, and Other Essays.* London: Longmans, Green.

Russell, B. (1940). *An Inquiry into Meaning and Truth.* London: George Allen and Unwin.

Russell, B. (1948). *Human Knowledge. Its Scope and Limits.* London: Allen and Unwin.

Ryle, G. (1954). *Dilemmas.* Cambridge University Press.

Sampson, G. (1978). Linguistic universals as evidence for empiricism. *J. Linguistics* **14**: 183–206.

Sánchez, F. (1581). *Que nada se sabe [De multum nobili et prima universali scientia quod nihil scitur].* Buenos Aires: Emecé, s.d.

Savage, L. J. (1972). *The Foundations of Statistics*, 2nd ed. New York: Dover.

Savage-Rumbaugh, E. S., D. M. Rumbaugh, and S. Boysen (1978). Symbolic communication between two chimpanzees. *Science* **201**: 641–644.

Schilpp. P. A., Ed. (1974). *The Philosophy of Karl R. Popper*, 2 vols. La Salle, Ill.: Open Court.

Schlick, M. (1925). *General Theory of Knowledge*. Transl. A. E. Blumberg. New York–Wien: Springer-Verlag, 1974.

Scribner, S. (1977). Modes of thinking and ways of speaking: culture and logic reconsidered. In Johnson-Laird and Wason, Eds. pp. 483–500.

Sebeok, T. A., and R. Rosenthal, Eds. (1981). *The Clever Hans Phenomenon: Communication With Horses, Whales, Apes, and People*. New York: New York Academy of Sciences.

Selfridge, O. G., and U. Neisser (1960). Pattern recognition by machine. *Sci. Amer.* **203** (Aug.): 60–68.

Shannon, C., and W. Weaver (1949). *The Mathematical Theory of Communication*. Urbana, Ill.: University of Illinois Press, 1963.

Sherif, M. (1936). *The Psychology of Social Norms*. New York: Harper and Bros.

Sherratt, A. (Ed.) (1980). *The Cambridge Encyclopedia of Archaeology*. New York: Cambridge University Press.

Shimony, A. (1971). Perception from an evolutionary point of view. *J. Phil.* **68**: 571–583.

Simon, H. A., and G. J. Groen (1973). Ramsey elimination and the testability of scientific theories. *Brit. J. Phil. Sci.* **24**: 367–380.

Simon, H. A. (1979). Information processing models of cognition. *Annual Rev. Psychol.* **30**: 363–396.

Skinner, B. F. (1948). Superstition in the pigeon. *J. Exp. Psychol.* **38**: 168–172.

Slobin, D. I. (1973). Cognitive prerequisites for the development of grammar. In C. A. Ferguson and D. I. Slobin, Eds., *Studies of Child Language Development*, New York: Holt, Rinehart and Winston.

Smedslund, J. (1963). The concept of correlation in adults. *Scandinavian. J. Psychol.* **4**: 165–173.

Smith-Churchland, P. (1978). Fodor on language learning. *Synthese* **38**: 149–159.

Smith-Churchland, P. (1980). A perspective on mind–brain research. *J. Phil.* **LXXVII**: 185–207.

Sneed, J. D. (1979). *The Logical Structure of Mathematical Physics*, 2nd ed. Dordrecht: Reidel.

Stegmüller, W. (1976). *The Structure and Dynamics of Theories*. New York: Springer-Verlag.

Stich, S. P. (1979). Between Chomskyan rationalism and Popperian empiricism. *Brit. J. Phil. Sci.* **30**: 329–347.

Suppe, F., Ed. (1974). *The Structure of Scientific Theories*. Urbana, Ill.: University of Illinois Press.

Suppes, P. (1970). *A Probabilistic Theory of Causality*. Acta Philosophica Fennica XXIV. Amsterdam: North-Holland.

Suppes, P., Ed. (1978). *Impact of Research on Education: Some Case Studies*. Washington, D.C.: National Academy of Sciences.

Swets, J. A. (1973). The relative operating characteristics in psychology. *Science* **182**: 990–1000.

Szentagothai, J. (1978). The neuron network of the cerebral cortex. *Proc. Roy. Soc. London* **B201**: 219–248.

Tarski, A. (1953). A general method in proofs of undecidability. In A. Tarski, A. Mostowski, and R. M. Robinson, *Undecidable Theories*. Amsterdam: North-Holland.

Tarski, A. (1956). *Logic, Semantics, Metamathematics*. Oxford: Clarendon Press.

Taylor, J. G. (1962). *The Behavioral Basis of Perception*. New Haven: Yale University Press.

Terrace, H. S., L. A. Pettito, R. J. Sanders, and T. G. Beaver (1979). Can an ape create a sentence? *Science* **206**: 891–909.

Teuber, H. L. (1965). Convergences, divergences, lacunae. In J. C. Eccles, Ed., *Brain and Conscious Experience*, pp. 575–583. New York: Springer-Verlag.

Thomas, R. D. K., and E. C. Olson (1980). *A Cold Look at Warm-Blooded Dinosaurs*. Boulder, Colorado: Westview Press.

Thompson, R. F. (1975). *Introduction to Physiological Psychology*. New York: Harper and Row.

Thompson, R. F., M. M. Patterson, and T. J. Teyler (1972). The neurophysiology of learning. *Annual Rev. Psychol.* **23**: 73–104.

Thurstone, L. L. (1930). The learning function. *J. General Psychology* **3**: 469–491.

Tolman, E. C., and I. Krechevsky (1933). Means-end-readiness and hypothesis. *Pychological Review* **40**: 60–70.

Törnebohm, H. (1979). Paradigms in Fields of Research. In I. Niiniluoto and R. Tuomela, Eds., *The Logic and Epistemology of Scientific Change*, pp. 62–90. Amsterdam: North-Holland.

Trehub, A. (1977). Neuronal models for cognitive processes: networks for learning, perception and imagination. *J. Theoret. Biology* **65**: 141–169.

Trudgill, Peter (1974). *Sociolinguistics*. Harmondsworth, Penguin Books.

Truesdell, C. (1974). A Simple example of an initial-value problem with any desired number of solutions. *Rendiconti dell'Istituto Lombardo*, Classe di Scienze (A) **108**: 301–304.

Truesdell, C. (1981). The role of mathematics in science as exemplified by the work of the Bernoullis and Euler. *Verhandl. Naturf. Ges. Basel* **91**: 5–22.

Truesdell, C. (1983). *An Idiot's Fugitive Essays on Science*. New York: Springer-Verlag.

Tuchman, B. (1978). *A Distant Mirror*. New York: Ballantine Books.

Tuomela, R. (1973). *Theoretical Concepts*. Wien-New York: Springer-Verlag.

Uexküll, J. von (1921). *Umwelt und Innenwelt der Tiere*, 2nd ed. Berlin: Springer-Verlag.

Umiker-Sebeok, J., and T. A. Sebeok (1981). Clever Hans and smart simians. *Anthropos* **76**: 89–165.

Unger, P. (1975). *Ignorance: A Case for Scepticism*. Oxford: Clarendon Press.

Vaihinger, H. (1920). *Die Philosophie des Als Ob*, 4th ed. Leipzig: Meiner.

Ville, J. (1939). *Etude critique de la notion de collectif*. Paris: Gauthier-Villars.

Vogt, E. Z., and R. Hyman (1959). *Water Witching USA*. Chicago: University of Chicago Press.

Volkmann, F. C., L. A. Riggs, and R. K. Moore (1980). Eyeblinks and visual suppression. *Science* **207**: 900–902.

Vollmer, G. (1975). *Evolutionäre Erkenntnistheorie*. Frankfurt: S. Hirzel.

Vygotskii, L. S. (1962). *Thought and Language*. Transl. E. Hanfman and G. Vakar. Cambridge, Mass.: MIT Press.

Walcott, C., J. L., Gould, and J. L. Kirschvink (1979). Pigeons have magnets. *Science* **205**: 1027–1029.

Ward, W. C., and H. M. Jenkins (1965). The display of information and the judgment of contingency. *Can. J. Psych.* **19**: 231–241.

Warnock, G. J. (1958). *English Philosophy Since 1900*. London: Oxford University Press.

Wason, P. C., and P. N. Johnson-Laird (1972). *Psychology of Reasoning*. London: Batsford.

Wasserman, G. S., G. Felstem, and G. E. Easland (1979). The psychophysical function: harmonizing Fechner and Stevens. *Science* **204**: 85–87.

Weiskrantz, L., and E. K. Warrington (1979). Conditioning in amnesia patients. *Neuropychologia* **17**: 187–193.

Wertheimer, M. (1961). *Productive Thinking*. London: Tavistock.

Westfall, R. S. (1973). Newton and the fudge factor. *Science* **179**: 751–758.

Whewell, W. (1847). *The Philosophy of the Inductive Sciences*, 2 vols. London: Frank Cass, 1967.

Whorf, B. L. (1956). *Language, Thought and Reality*. New York: Wiley.

Wiesel, T. N., and D. H. Hubel (1965). Comparison of the effects of unilateral and bilateral eye closure on cortical unit responses in kittens. *J. Neurophysiol.* **28**: 1029–1040.

Wilder, R. (1981). *Mathematics as a Cultural System*. Oxford: Pergamon.

Williams, B. A. O. (1972). Knowledge and reasons. In G. H. v. Wright, Ed., *Problems in the Theory of Knowledge*. The Hague: Martinus Nijhoff.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.

Wolpe, J. (1976). How laboratory-derived principles of learning have conquered the neuroses. In G. Serban, Ed. *Psychopathology of Human Adaptation*. New York: Plenum.

Wolpe, J. (1978). Cognition and causation in human behavior and its therapy. *Amer. Psychologist* **33**: 437–446.

Woodruff, G., D. Premack, and K. Kennel (1978). Conservation of liquid and solid quantity by the chimpanzee. *Science* **202**: 991–994.

Woodruff, G., and D. Premack (1981). Primative mathematical concepts in the chimpanzee: proportionality and numerosity. *Nature* **293**: 568–570.

Ziman, J. (1968). *Public Knowledge*. Cambridge: Cambridge University Press.

Ziman, J. (1979). *Reliable Knowledge*. New York: Cambridge University Press.

INDEX OF NAMES

Abelson, P. 110
Ackoff, R. 239, 246, 285
Adams, J. L. 239
Agassi, J. 224, 227
Albert, M. L. 13, 34
Alcock, J. 37
Amari, S.-I. 26
Andreski, S. 614
Aquinas, T. 214, 373
Archimedes 373
Aristotle 5, 77, 373
Asch, S. E. 82
Austin, G. 287
Avenarius, R. 99, 152, 304
Ayer, J. A. 152

Bach, J. S. 244
Bacon, F. 67, 134, 137, 186, 257, 373
Bandura, A. 13
Baranyi, A. 25
Barber, T. X. 82, 138
Bartlett, F. C. 38, 132, 136
Battro, A. M. 149, 150
Beaulieu, A. 111
Bechtereva, N. P. 40
Becker, G. 105
Beer, C. G. 141
Békesy, G. von 7, 131, 139
Bell, J. L. 337, 373
Bellugi, U. 114
Berger, P. L. 156
Bergson, H. 174
Berkeley, G. 151
Berlyne, D. 236
Bindra, D. 7, 8, 10, 22, 25, 35, 38, 41, 46, 116, 262
Birch, H. G. 238, 248
Black, M. 214
Bliss, T. V. P. 25
Bloor, D. 121
Bohr, N. 156
Boltzmann, L. 354
Bolzano, B. 23, 65
Bondi, H. 239
Botkin, J. W. 112
Bouchard, T. J. Jr 55
Brewer, D. 106
Bridgman, P. W. 152
Brines, M. L. 114
Brown, H. I. 137
Bruner, J. S. 287
Bunge, M. 8, 22, 27, 37, 39, 41, 58, 60, 71, 110, 112, 156, 163, 165, 196, 211, 249, 265, 269, 311, 348, 365, 372

Campbell, D. T. 58
Carnap, R. 99, 152, 349, 374
Carterette, E. C. 135
Chapanis, A. and N. P. 201
Chapman, J. P. and J. L. 202
Čapek, M. 8
Chomsky, N. 44–45, 79, 97, 119–120, 182, 303
Claparède, E. 287
Clark, E. V. 121
Cohen, N. L. 78
Coleman, J. 243
Comte, A. 93, 123, 219
Craig, W. 374
Currie, G. 59

Darwin, C. 7, 8, 58, 289, 319
Davenport, L. D. 85
Dawson, M. E. 46
de Finetti, B. 309
Denison, E. F. 105
Descartes, R 4, 149, 193, 373
Dickinson, A. 46, 63
Dodwell, P. 54

Dreistadt, R. 248
Duhem, P. 5, 140
Duncker, K. 235, 238
Dyson, F. 320

Eccles, J. C. 8
Edison, T. A. 111, 250
Einhorn, H. J. 47–48
Einstein, A. 160–161, 354
Ellis, A. E. 138
Eliot, T. S. 66
Euclid 373
Euler, L. 77

Fehér, O. 25
Feller, W. 306
Ferrater-Mora, J. M. 121
Feigenbaum, E. A. 296
Festinger, L. 201
Feyerabend, P. K. 186, 224
Fleck, L. 105, 156, 301
Flohr, H. 25
Fodor, J. 27, 121
Freud, S. 182
Friedman, M. P. 135
Furedy, J. J. 46

Galilei, G. 227
Garcia, J. 223
Gardner, B. T. 115
Gardner, R. A. 115
Gaviola, E. 246
Gibson, J. 34, 132
Gilbert, C. D. 28
Glass, L. 190, 312
Goddard, G. V. 25
Goodman, N. 152
Goodnow, J. J. 287
Goody, J. 123, 187
Gould, J. L. 114
Gould, S. J. 145
Greenberg, J. H. 133
Greenwell, J. R. 84
Groen, G. J. 374
Gruber, H. 249

Hájek, P. 295
Hansel, C. E. M. 37
Hanson, N. R. 140, 295
Harnad, S. 117
Hávranek, T. 295
Hebb, D. O. 7, 8, 22, 25–26, 32, 35, 38, 39, 40, 120, 248
Hécaen, H. 13, 34
Hegel, G. W. F. 65, 124, 210
Heidegger, M. xv, 615
Hein, A. 32
Held, R. 32
Hellman, R. H. G. 312
Helmholtz, H. von 7, 58, 136, 193
Hempel, C. G. 365
Hennessy, J. W. 235
Herbart, J. F. 190
Hesse, M. 214
Hessen, B. 107, 242
Hilbert, D. 5, 245, 374
Hinde, R. A. 140
Hintikka, J. 81, 88
Hippocrates 4
Hobbes, T. 10
Hoffman, W. C. 136
Hogarth, R. M. 47–48
Holton, G. 145, 268, 349
Hubel, D. H. 28, 53, 135
Hume, D. 160
Hunter, W. S. 35
Husserl, E. 94
Hyman, R. 249

Inhelder, B. 150

James, W. 77, 120, 153, 157
Jantsch, E. 261
Jastrow, R. 391
Jenkins, H. M. 47, 310
Jerison, H. J. 58
Johnson-Laird, P. N. 203

Kant, I. 5, 94, 121, 134, 137, 149, 151, 152, 160, 197, 214, 373
Kapitza, P. 263
Keynes, J. M. 107, 108
Katona, G. 238
Kendrew, J. 244

Kierkegaard, S. 86
Klima, E. S. 114
Köhler, W. 7
Krechevsky, I. 287
Kuhn, T. S. 107, 129, 156, 224–225
Kyburg, H. J. Jr 310

Lambert, W. E. 119
Lamarck, J. B. 58
Lashley, K. 287
Latour, B. 10, 156
Lazarsfeld, P. 243
Lederberg, J. 296
Lehrer, K. 86
Leibniz, G. W. von 4, 94, 95, 295, 373
Lenin, V. I. 193
Lenzen, W. 88
Leonardo da Vinci 73, 130, 139, 152
Le Roy, E. 129
Leuba, C. 236
Levine, M. 287
Levine, S. 235
Lewis, C. I. 77
Lewis, W. 105
Lieberman, P. 117
Lindauer, M. 28
Linnaeus, C. von 373
Lloyd, G. E. R. 187
Lloyd Morgan, C. 7
Locke, J. 157
Loehlin, J. C. 55
Lorenz, K. 7
Lovejoy, A. O. 98, 134
Luckman, T. 156
Luneburg, R. K. 149
Luria, A. R. 104, 121, 200

Mach, E. 7, 8, 58, 99, 131, 152, 157, 304, 365, 374
Mackey, M. 312
Mackintosh, N. J. 46
Magoun, H. W. 235
Marx, K. 107
Masterton, R. B. 58
Maudsley, H. 22, 56
Mayer, R. E. 202, 287
Maxwell, J. C. 290
McCloskey, M. 67
McGue, M. 55
McKenna, M. C. 292
Medawar, P. 15
Meier, N. R. F. 240
Mendel, G. 145
Mermin, N. D. 156
Merton, R. K. 14, 107, 223
Michelangelo 152
Mill, J. S. 99, 152, 153
Miller, G. A. 49, 50, 131
Millikan, R. 145
Milner, P. 7
Mountcastle, V. 25
Minsky, M. 237
Modigliani, F. 313
Moruzzi, G. 235

Nadel, L. 148–149
Nader, R. 225
Nagel, E. 365
Neisser, U. 50, 132, 135, 136, 147, 202
Nelson, L. xiii
Newell, A. 263
Newton, I. 145, 245, 373
Nichols, R. C. 55
Nisbett, R. xiv, 200, 203
Nietzsche, F. 7, 8, 92, 214
Noton, D. 147
Nottebohm, F. 64

Oakeshott, M. 77
O'Keefe, J. 148–149
Olds, J. 32
Olson, E. C. 320

Park, R. 243
Parmenides 174
Pasteur, L. 300
Patterson, F. 115
Pauling, L. 196
Pavlov, I. P. 233
Peccei, A. 49
Peirce, C. S. 58
Penfield, W. 32
Phelps, M. 133

Piaget, J. 7, 46, 58, 104, 118, 121, 132, 136, 149, 150, 163–164, 200–201
Pincus, J. H. 13, 34
Plato 4, 65, 76, 77, 92, 94, 118, 124, 152
Poincaré, H. 224
Polanyi, M. 77, 268
Pólya, G. 265
Popper, K. R. 8, 23, 47, 58, 65, 69, 88, 94, 124, 223, 224, 227, 361, 363, 368
Precht, W. 25
Premack, D. 115
Protagoras 152
Ptolemy 4
Putnam, H. 23, 27
Pylyshyn, Z. 27

Quine, W. V. O. 118, 344

Rabinowitz, H. S. 238
Ramón y Cajal, S. 25, 73, 290
Ramsey, F. P. 374
Randi, J. 249
Rasiowa, H. 373
Ratliff, F. 130
Reichenbach, H. 349
Rescher, N. 15, 206
Restle, F. 287–288
Rignano, E. 228
Roche, M. 105
Romer, A. S. 292
Rorty, R. xiii, 67
Rosch, E. 122
Ross, L. xiv, 200, 203
Russell, B. 77, 86, 99, 152, 221, 321
Ryley, G. 221

Sampson, G. 119
Sánchez, F. 92
Savage, L. J. 309
Savage-Rumbaugh, E. S. 115
Schlick, M. 197, 208
Scribner, S. 202
Sebeok, T. A. 112, 115
Selfridge, O. G. 135
Sextus Empiricus 92
Shannon, C. 34, 103
Sherif, M. 82
Sherratt, A. 261
Sikorski, R. 373
Simon, H.A. 263, 374
Skinner, B. F. 85
Smedslund, J. 289, 310
Sneed, J. 337, 355
Socrates 182, 251
Spence, K. 287
Spencer, H. 7, 8, 58, 93
Squire, L. R. 78
Stark, L. 147
Stegmüller, W. 337, 355
Stich, S. P. 119
Slobin, D. I. 121
Smith-Churchland, P. 121
Suppe, F. 338
Suppes, P. 13, 190
Swets, J. A. 145
Szentágothai, J. 25

Tarski, A. 348, 373
Taylor, J. G. 134
Terrace, H. S. 115
Thomas, R. D. K. 320
Thompson, R. F. 150
Tolstoy, L. 140
Tolman, E. C. 287
Trudgill, P. 122
Truesdell, C. 312, 337, 339
Tucker, G. J. 13, 34
Takeuchi, A. 26
Teuber, H.-L. 242
Tuchman, B. 201
Tuomela, R. 375

Uexküll, J. von 130, 148
Unger, P. 92
Umiker-Sebeok, J. 115

Vaihinger, H. 214
Vesalius 73
Ville, J. 307
Vogt, E. Z. 249
Volkmann, F. C. 133
Vollmer, G. 8, 59

Wald, G. 242
Ward, W. C. 47, 310

Warnock, G. J. 283
Warrington, E. 39
Wason, P. C. 47, 203
Watson, J. B. 152
Weaver, W. 34, 103
Weiskrantz, L. 39
Wertheimer, M. 238
Whewell, W. 4, 134, 137, 140, 296, 373
Whitehead, A. N. 357
Whorf, B. L. 121
Wiesel, T. N. 28, 53, 135
Wilder, R. 70
Wilson, S. C. 138
Wittgenstein, L. XV, 97, 122, 123, 187, 221, 244, 283
Wolpe, J. 13
Woodruff, G. 150
Woolgar, S. 10, 156
Wundt, W. 37

Yerkes, R. M. 7

Ziman, J. 107, 145
Zinnes, J. L. 190

INDEX OF SUBJECTS

Abstraction 172–173
Activity of a thing 30–32
Agnosticism 93
Agreement, degree of 207–208
Analogism 214
Analogy 210–213
Analysis 75, 219–22, 274
Ape language 115–117
Appearance 150–154
Apprenticeship 49–50
A posteriori 181–182
A priori 181–182
Approach 258–262
Artifact 65
Ars inveniendi 295
Artificial intelligence 27
Association 178, 305
Assumption 341, 359. *See also* Hypothesis
Atomism 165, 259–260
Attribute 165–167, 187–191
Attribution 177
Axiomatics 348–349

Background 332–333
Bayesian epistemology 71–72, 357
Belief 80–88, 91
Biology 6–8
Black-box model 301–302

Chunking 49
Classification 324–331
Clever Hans phenomenon 115–116
Cognition 61ff.
Cognitivism 9–10, 27
Cognitology 11–12
Collecting 169–171
Committee 111
Communication 97–125
Compatibility, degree of 208
Computation 338–339
Computer 95–96, 263, 295, 339
Conceiving 159–198
Concept 159–174, 185–186, 340, 374
Conditioning 179
Conjecture 286–322
Construct 41. *See also* Concept, Proposition, and Theory
Consistency 206–207
Context 323
Contradiction 206–207
Contrast enhancement 130–131
Controversy 224–228
Convention 355–359
Correlation, statistical 306
Counterfactual 176
Craft 92
Creativity 195–198
Criticism 222–228

Data 145, 342–344
Deduction 203–206
Deductivism. *See* Refutationism
Deficit, cognitive 33
Definition 340–341, 356–358
  operational 335. *See also* Indicator
Denotation 356
Depth, conceptual 300–303
Designation 356
Detection 143
Development, cognitive 50–57
Devil's trident 155
Diagram 192
Dialectics 68–70, 165
Discovery 194–195
Discrimination 164–165
Dissonance, cognitive 201
Divination 249
Dogmatism 85–86. *See also* Ideology and Religion
Doubt 81

Dualism, mind-body 9, 65, 98, 147
Duality postulate 98

Economy of thought 304–305
Einstellung 139, 238
Empathy 138
Empiricism 94, 102, 119, 157, 197, 228–229, 294
Equivalence 168, 171
  functional 63
  relation 326–327
ESP 139
Ethics 12
Event 29
Evolution, biological 377–379
  cognitive 57–59, 67–72
Evolutionism, epistemological 8, 58–59
Exactification 185–186
Existence criterion 154
  problem 275–276
Exploration 233–285
Extension of a predicate 166, 171, 183–184

Fact-finding 236
Faith 282
Fallacy 202–203
Fallibilism 93
Fashion 244
Fictionism 214
Field of knowledge 89–92
Filter, sensory 130–131
Formalism 355
Framework 90
Frequency, relative 307
Functionalism 27

Generalization 178–180, 361–362, 366–368
Geometry 149–150
Gödel's incompleteness theorem 281
Grammar 115–116, 119–120, 303
Group belief 83
Group learning 112
Guess, educated 288
Gullibility 84–85

Habitat, cognitive 103–104
Hermeneutics 123, 221–222
Heuristics 265–269
Holism 165, 260
Hypothesis 217, 288–323
  ad hoc 318–319, 352
  driving 140–141
Hypothetico-deductive method 209

Ideafication 167
Idealism 94, 102
Ideology 394–395
Ignorabilia 93
Imagination 138
Inconceivable 174
Indicator 144, 217, 335
Induction 214–216
Inductivism 229
Inference 199–229
Information 34, 103, 132, 213
Inhibition 131–132
Innatism 44–45, 54
Innovation 248
Inquiry 89–91, 241–242
Instrument 143–144
Inter-subject relation 101–103
Intuition 217, 248–250
Invention 195

Judgment 175–176
Justification 86–87

Know-how 77–78
Know-that 77–78
Knowledge 61–96
  field of 89–91
  industry xiii, 92

Language 40, 114–125
  acquisition 79, 116–120
Law 359–373
Learning 24–50, 62–67
  community 108–112
  theory 43–46
Literacy 123–124
Logic 12, 204–206, 208, 316

Magnitude 189

Map, cognitive 148
  conceptual 191–196
  perceptual 146–150, 192–193
Mapping of mental functions 32, 37–38
Materialism 9, 23, 99–100
Mathematics, role of in theory construction 345–346
Measurement 190, 343–344
Mechanism 303–304
Mentality, primitive 202
Memory 38–39
Metanomological statement 372
Meteorology 354
Method 250–258
Methodics 259
Methodoclasm 15, 257
Methodolatry 15, 257
Model, theoretical 335–337, 351
Monetarism 313
Monte Carlo method 366
Motivation 46, 240

Nature-nurture 54–56
Negation 166
Neurolinguistics 118
Neuropsychology 8–10, 56, 94
Niche, cognitive 103–104
Nominalism 167

Objectivity 80, 155–156
Observation 139–146
Ockham's razor 304
Operationism 321, 355, 374
Opinion 76
  prevailing 83–84
Originality 42, 66, 195

Partition of a set 171, 324–330
Pattern 79, 359–360
Perception 35–37, 129–157
Pexgo 28. *See also* Psychon
Phenomenalism 151–153
Phenomenon 150–154
Plasticity, neuronal 24–26, 28, 62–64, 161–162
Positivism 163
Postulate 341
Precision 316–317
Probability 71–72, 307–310, 365
Problem 233–285
Problematics 259, 576
Process 30–32
Proof 205
Property 165–168, 187–191, 329
Proposition 175–177, 180–181
Psychon 24, 63
Pythagoreanism, semantic 355

Quantitation 188–190
Quantum mechanics 314
Rationalism 197, 229
Realism 94, 102, 158, 197, 354, 375, 376
Reality 151–152
Reason 204
Reasoning 200–218
Reference 183–184
Reification 167
Representation 191–194
Research 110–111, 262–269
Rule 79, 369–373

Schema, perceptual 136–137
Science, cognitive 9
Scientometrics 72
Semantic assumption 350–355, 380–381
Sensation 35–36, 130–133
Sense organ 130–131
Sequential machine 265
Set 170–171, 186. *See also Einstellung*
Similarity 169
  degree of 212
Simplicism 304–305
Simplification 346, 359
Skepticism 85, 92–93, 225. *See also* Fallibilism
Social conditions of inquiry 104–108
Sociologism 106–108
Space, cognitive 41–42
  public 104
Species 328
Speculation 293, 320
State 29
  space 29–30
Statistics, mathematical 217–218

Strength, logical 315–316
Subject-object relation 97–100
Subjectivism 156–157
Substitution 204
Subtheory 334
Superstition 84–85
Synthesis 274
System, social 109
System, representation of 192, 301–302
Systematization 323–376. *See also* Classification and Theory

Taxonomy. *See* Classification
Technique 251–253
Theory 331–355, 375–376
  general 335–337
  kinematical 334
  mechanismic 334
  specific 335–337. *See also* Model, theoretical
Thought 39, 42, 120–121, 174–176, 187
Thouless index 150–151
Tree questioning 251
Trend 360–361
Truth 205–206, 254
Turing machine 27, 370

Uncertainty 85
Unit 358–359

Vagueness 182–185

Whorf-Sapir thesis 121–122

Mario Bunge

Treatise on Basic Philosophy

Volume 6

SPRINGER-SCIENCE+BUSINESS MEDIA,B.V

TREATISE ON BASIC PHILOSOPHY

Volume 6

EPISTEMOLOGY AND METHODOLOGY II: UNDERSTANDING THE WORLD

TREATISE ON BASIC PHILOSOPHY

	1
SEMANTICS I	*Sense and Reference*
	2
SEMANTICS II	*Interpretation and Truth*
	3
ONTOLOGY I	*The Furniture of the World*
	4
ONTOLOGY II	*A World of Systems*
	5
EPISTEMOLOGY & METHODOLOGY I	*Exploring the World*
	6
EPISTEMOLOGY & METHODOLOGY II	*Understanding the World*
	7
EPISTEMOLOGY & METHODOLOGY III	*Philosophy of Science & Technology*
	8
ETHICS	*The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 6

Epistemology & Methodology II:

UNDERSTANDING THE WORLD

SPRINGER-SCIENCE+BUSINESS MEDIA, B.V

Library of Congress Cataloging in Publication Data

CIP

Bunge, Mario Augusto.
Understanding the world.

(Epistemology & methodology ; 2) (Treatise on basic philosophy ; v. 6)
Bibliography: p.
Includes indexes.
1. Knowledge, Theory of. I. Title. II. Series: Bunge, Mario Augusto. Epistemology & methodology ; 2. III. Series: Bunge, Mario Augusto. Treatise on basic philosophy ; v. 6.
BD161.B86 1983 no. 2 121 s [121] 83-14001
ISBN 978-90-277-1635-4 ISBN 978-94-015-6921-7 (eBook)
DOI 10.1007/978-94-015-6921-7

All Rights Reserved
© 1983 by Springer Science+Business Media Dordrecht
Originally published by D. Reidel Publishing Company, Dordrecht, Holland in 1983
No part of the material protected by this copyright notice may be reproduced or utilized in any form or by any means, electronic or mechanical, including photocopying, recording or by any information storage and retrieval system, without written permission from the copyright owner

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and of right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient—except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can

become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: "Do your own thing. Your reward will be doing it, your punishment having done it".

CONTENTS OF EPISTEMOLOGY II

GENERAL PREFACE TO THE *TREATISE* v

PREFACE TO *EPISTEMOLOGY II* xi

SPECIAL SYMBOLS xiii

PART IV. UNDERSTANDING AND CHECKING 1

10. UNDERSTANDING 3
  1. Understanding and Explaining 4
    1.1. Modes of Understanding 4
    1.2. Subsuming and Explaining 8
  2. Systematic Account 16
    2.1. Vulgar, Ideological, and Scientific Accounts 16
    2.2. Basic Explanation Types 24
  3. Unification 31
    3.1. Reduction 31
    3.2. Integration 41
  4. Forecasting 45
    4.1. From Expectancy to Forecast 45
    4.2. Scientific Prediction 52
  5. Concluding Remarks 58

11. PRODUCING EVIDENCE 59
  1. From Self-Evidence to Evidence 61
    1.1 Self-Evidence 61
    1.2 Evidence 65
  2. Testability and Indicators 72
    2.1 Testability 72
    2.2. Indicators 85

3. Data 91
  3.1. Measurement 91
  3.2. Experiment 102

4. Concluding Remarks 111

12. EVALUATING 114
1. Values 115
  1.1 Truth 115
  1.2. Usefulness 129

2. Empirical Value Indicators 132
  2.1. Confirmation 132
  2.2. Efficiency 140

3. Conceptual Value Indicators 143
  3.1. External Consistency 143
  3.2. Other 148

4. Concluding Remarks 153

PART V. VARIETY AND UNITY 155

13. EPISTEMIC CHANGE 157
1. Cognitive Novelty 158
  1.1. Cognitive Kinematics 158
  1.2. Conceptual Change 162

2. Change Mechanisms 165
  2.1. Usual Mechanisms 165
  2.2. Reduction and Synthesis 171

3. Evolution and Revolution 175
  3.1. Paradigm 175
  3.2. Revolution 179

4. Limits and Prospects 184
  4.1. Limits to Inquiry 184
  4.2. Prospects of Inquiry 189

5. Concluding Remarks 191

14. KINDS OF KNOWLEDGE 194
1. Fields of Knowledge 195
  1.1. Knowledge Genera 195
  1.2. Research Field 197

2. Science and Technology 200
  2.1. Basic Science 200
  2.2. Applied Science and Technology 207

3. The Knowledge System 215
  3.1. Inter-Relations 215
  3.2. Mergers and Mixtures 219

4. Illusory Knowledge 223
  4.1. Pseudoscience 223
  4.2. Ideology 228

5. Concluding Remarks 237

15. UPSHOT 240
1. Social Sciences of Knowledge 241
  1.1. Descriptive Social Sciences of Knowledge 241
  1.2. Inquiry Policies 247

2. Philosophies of Knowledge 252
  2.1. Rationalism and Empiricism 252
  2.2. Realism and Scientism 258

3. Maxims of Scientific Realism 264
  3.1. Descriptive Principles 264
  3.2. Regulative Principles 267

4. Concluding Remarks 270

APPENDICES 272

3. Partial Truth 272

4. Predictive Power 276

5. Formal Structure of Experiment 278

6. Degree of Confirmation of a Theory 281

BIBLIOGRAPHY 283

INDEX OF NAMES 291

INDEX OF SUBJECTS 294

PREFACE TO *EPISTEMOLOGY II*

This is the sequel to *Epistemology I: Exploring the World.* In that work we studied cognition as a brain process, and communication as a social transaction. In particular, we studied perception and conception, the formation of propositions and proposals, exploration and systematization, discovery and invention. We regarded knowledge as an outcome of processes occurring in animals that learn by themselves and from one another. We took concepts and propositions, problems and proposals, to be equivalence classes of brain processes rather than ideal objects detached from brains and from society. However, we also stressed the need for studying such abstractions as well as the corresponding real processes.

In other words, we admitted that cognition ought to be studied both concretely (as a biopsychosocial process) and abstractly (with disregard for personal and social idiosyncrasies). We hoped in this way to favor the merger of the various hitherto separate approaches to the study of knowledge and knowledge-directed action: the neurophysiological and the psychological, the sociological and the historical, the epistemological and the methodological ones. After all, these various approaches have a single aim, namely to improve our understanding of the ways we get to know reality, and the ways knowledge can be utilized to alter reality.

In this volume we will study the ways theories and proposals (e.g. technological designs) are put to the test and used to understand or alter reality. We will stress the difference between belief and inquiry. We will study the kinds of knowledge and the ways human knowledge grows, declines, or alters course. We will distinguish basic science from applied science, and both from technology and ideology, and we will seek to demarcate genuine knowledge from bogus. We will analyze the two mechanisms for enhancing the cross-fertilization and the unity of the various branches of knowledge: reduction and integration. We will stipulate the conceptual and empirical conditions a proposition has to fulfill in order to be valued as (sufficiently) true, and a proposal to be regarded as (suitably) efficient. (We shall do so in the light of real cases drawn from contemporary research rather than in obedience to a priori philosophical principles.) We will analyze a number of important yet vague notions, such as those of truth and efficiency, background and framework, paradigm and revolution. And we will explore the possible limits to our exploration of the world, as well as the limitations of the classical philosophies of knowledge.

The upshot of our study is a descriptive and normative epistemology that cannot be compressed into a couple of slogans, although it combines some features of rationalism with others of empiricism. This synthesis may be called *scientific realism* because the criterion for adopting or rejecting any given thesis is its compatibility or incompatibility with the practice of research in contemporary science (basic or applied), technology, or the humanities. We find no use for a theory of knowledge, however exact or ingenious it may be, that is divorced from knowledge.

SPECIAL SYMBOLS

$\neg p$	*not-p*
$p \& q$	*p and q* (*conjunction*)
$p \vee q$	*p or q* (*disjunction*)
$p \Rightarrow q$	If *p*, then *q* (*implication*)
$p \Leftrightarrow q$	If *p*, then *q*, and conversely (*iff*)
$A \vdash B$	Premises *A entail* conclusions *B*
$=_{df}$	Identical by *definition*
$\{x \mid Fx\}$	The *set* of objects possessing property *F*
$\emptyset$	The *empty set*
$a \in A$	Individual *a belongs to* set *A*
$A \subseteq B$	Set *A* is *included* in set *B*
$A \cup B$	The sets of objects in *A or* in *B*
$A \cap B$	The set of objects in *A and* in *B*
$A - B$	The set of objects in *A* but *not in B*
$A \Delta B$	The set of objects in *A or* in *B* but *not in both*
$\langle a, b \rangle$	The *order pair* of *a* and *b*
$A \times B$	The *cartesian product of A* and *B*, i.e. the set of ordered pairs $\langle a, b \rangle$, where *a* is in *A* and *b* in *B*
2^A	The *power set* family of all subsets) of set *A*
$\lvert A \rvert$	The *cardinality* (numerosity) of set *A*
$\mathbb{N}$	The set of *natural numbers* (0, 1, 2, . . .)
$\mathbb{R}$	The *real line*
$\mathbb{R}^+$	The set of *non-negative real numbers*
$f: A \rightarrow B$	The *function f* maps set *A* into set *B*
$f(x)$	The *value* of function *f* at *x*
$F = \langle F_1, F_2, \ldots, F_n, \ldots \rangle$	A state function for some system
R & D	Research and development
S & T	Science and technology

PART IV

UNDERSTANDING AND CHECKING

CHAPTER 10

UNDERSTANDING

We do a number of things, from eating to learning to write, without understanding what we are doing. Automatic machines and computers, too, perform their task without understanding. But their designers, repairmen and users are supposed to understand, at least to some extent, what they do. All intellectual and technical operations require some understanding unless they follow routines; and even when they do so the control of the results requires some understanding of the nature, goal and value of the operations.

Understanding is not an all-or-none operation: it comes in several kinds and degrees. Even after studying a subject for decades we may feel that we can still make progress at understanding it: we learn to see it in different ways, to relate it to different themes, or to use it in different ways. Until a time may come when, having understood the subject fairly well, we realize that it is error ridden or has become obsolete.

Consider the following imaginary dialogue. "What is new?—President *X* died.—How, when, where?—He was murdered yesterday at *Y*.—Why?—The inquiry is still under way." The point of this dialogue is to note three different and successive stages in any thorough inquiry into factual matters, namely finding the *what*, the *how*, and the *why*—or information, description, and explanation respectively.

A set of raw data tells us that something is or is not the case. A description tells us how the fact has come about. An explanation is one more step in the process of inquiry. It involves conjecturing mechanisms of some kind—causal, stochastic, teleological, or a combination of these—underlying the facts being described. In the above example any of the following may turn out to be part of the correct explanation. The assassin was crazy; he wanted to attain notoriety; he was an enemy agent; he had a grudge to avenge; he regarded the victim as public vermin; he wanted to replace the victim; he wanted to trigger an uprising. Any of these statements, though schematic, goes beyond the mere description of the fact: it is a hypothesis. By adopting any of it we may expedite or block the research and thus favor or hinder the process of refining the original description.

We may understand or fail to understand the what, the how, and the why of a fact, and we may do so in different manners and to different degrees. But, in all cases of genuine knowledge, knowing that something is (or could be) the case precedes knowing how, which in turn precedes knowing why. (This does not hold for ideological explanation, which dispenses with inquiry altogether.) It follows that explanatory knowledge is superior to descriptive knowledge, which is in turn superior to mere information. That is, knowing why implies knowing how, which in turn implies knowing that.

To admit the above involves adopting an epistemology at variance with empiricism, which counsels description and discourages explanation. By the same token it involves sharing the belief, widespread among scientists, technologists, and humanists, that explanation is a goal of research as distinct from mere inquiry, which usually stops as the how. But admitting all this is of little value unless we know what is explanation. We shall therefore undertake to examine explanation.

1. UNDERSTANDING AND EXPLAINING

1.1. *Modes of Understanding*

We can understand, or misunderstand, objects of various categories: facts, signs, and constructs. Thus we can understand what happens around us, we can understand a sentence, and we can understand a hypothesis. In principle anything can be the object of understanding—even non-facts such as travel at superluminal speed and the adventures of disembodied spirits. Nor are the understandings of objects of different types unrelated. Thus to understand a fact one needs to know other facts and some hypotheses. And to understand a theory one needs to understand some of the sentences in which its formulas are expressed.

One may understand in different ways and degrees. One may understand with the help of words or diagrams, examples or analogies, empathy or hypotheses. To be sure, these various modes of understanding are inequivalent, but the point is that they constitute understanding even if not all of them amount to genuine explanation. Thus we may understand the apparently odd behavior of someone else if we make an effort to put ourselves in his shoes: this is an act of empathic understanding or *Verstehen*, made possible by the basic similarity of all human brains. We understand *A*'s dislike for *B*'s philosophy when informed that B stole A's bride (or hypothesis or experiment). Dalton understood atomic combinations in terms of tiny hooks. Faraday understood electromagnetic induction in terms of elastic tubes. Darwin understood natural selection by likening it to economic competition.

Metaphorical explanation is not limited to lay discourse: one often finds it in the beginning stages of scientific and technological developments, even in sophisticated fields such as many body physics. Thus consider a network of atoms or molecules. One of them absorbs a photon and so gets excited, i.e. jumps to a higher energy state. This excitation, an event, is reified into an "exciton", which is treated *as if* it were a particle. Indeed, when the excitation propagates to a neighboring atom or molecule, the exciton is said to pass on from one atom (or molecule) to the next. Since this hopping goes on at random, it is treated *as if* the exciton "were enjoying a random walk through the lattice", as one author puts it. (I.e., the metaphorical exciton is likened to a drunkard—a metaphor of a metaphor—so that the theory of random walk can be used.) Eventually the exciton ceases to wander: this event is treated *as if* the exciton had fallen into a trap ("trapping center") where the exciton triggers a chemical reaction. The entire metaphor has as much heuristic value as Faraday's analogy between the electromagnetic field and the elastic solid. And in both cases this heuristic scaffolding is eventually dismantled: only the equations remain, which are interpreted in a different (literal) way. A new, superior, mode of understanding is attained.

In any of its modes understanding involves systematizing—or, as Piaget would have said, assimilating. Either we fit the given item into our pre-existing cognitive or epistemic network, or we transform (e.g. expand) the latter to accommodate the new item. See Figure 10.1. Needless to say, this is a metaphorical explanation of the phenomenon of understanding. Like some other metaphorical explanations, it may prove to be of heuristic value: it may suggest a precise formulation in terms of changes in the state space of a psychon or plastic neural system.

Every student is familiar with the phenomenon of the growth of understanding, which at times is gradual and at other times sudden, and at all times is a matter of degree or level. The first reading of a paper may give him a dim understanding, a second a deeper understanding, and a third, perhaps apparently unrelated experience, may show how everything falls into place. We understand a thing or a process if we are given a description of it; we understand it even better if the description follows from some mechanismic model.

What holds for individual learning holds, *mutatis mutandis*, for the social

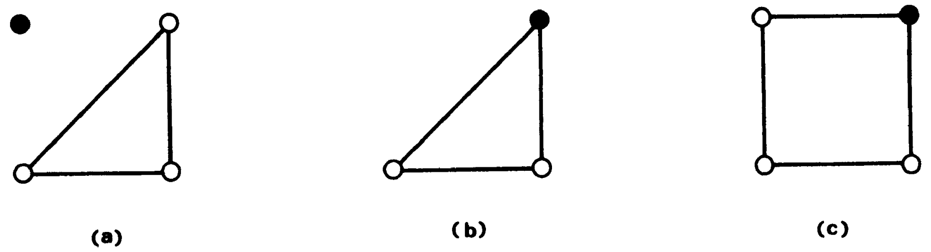


Fig. 10.1. Understanding as systematization. (*a*) Before understanding item symbolized by black dot. (*b*) Item is understood by fitting into pre-existing epistemic network. (*c*) Item is understood by transforming epistemic network.

process of growth and diffusion of knowledge. Along the history of any epistemic field metaphors are replaced by literal descriptions, which are systematized into black boxes, which are superseded by mechanismic theories, which in turn make it possible to refine descriptions. A few historical examples will help us understand this point.

Take for instance solidity: we understand it on being told that the inter-particle distances in a solid are constant. If we now ask why those distances are constant, we demand a theory involving inter-particle forces, i.e. a mechanismic or dynamical theory. Take disintegration next: an exponential decay curve, or the corresponding equation, answers the how-question but does not explain why there is disintegration rather than stability. An explanation requires a theory involving the hypothesis that special forces are involved. Or take high energy processes such as the high speed collision of protons with neutrons and the resulting ménagerie of particles. Until recently physicists were satisfied with data fitting and with black box models capable of absorbing and occasionally predicting large amounts of data but incapable of explaining them. The currently popular theories go far beyond that phenomenology and hypothesize that many a so called "observed particle" is a system—e.g. one composed of quarks and antiquarks bound together by peculiar forces.

Nor is the above pattern of progress restricted to physics. Chemistry has long ceased to be a huge catalogue of reactions and compounds. We know now the composition and structure of millions of compounds, and are beginning to explain their very existence in terms of forces and collisions. Likewise biochemistry is advancing from description to explanation. At the time of Claude Bernard biochemistry was mainly concerned with establishing balance sheets, i.e. with relating what the organism consumes with what it excretes. For example, the amount of protein consumed was correlated with the amount of urea and creatinin in urine: the intermediate processes were unknown. Nowadays a number of metabolic pathways have been uncovered, and many biosynthesis processes, such as photosynthesis, have been analyzed. The same holds for the adjoining field of genetics. Classical genetics treated the organism as a black box and, accordingly, performed observations and experiments on the whole organism level—e.g. hybridization experiments. With the discovery and chemical identification of the gene, and particularly with the formulation and confirmation of the double helix hypothesis, the genome became a grey box. True, biochemistry and genetics are still primarily structural and kinetic, and only marginally dynamical. But it is likely that more and more dynamical (mechanismic) models, explaining the known structures and processes, will be proposed as theorists invade the field in increasing numbers. As Francis Crick, of double helix fame, once said, "One man's black box is another man's problem".

Biology is no exception to the general trend from description to explanation. True, much of biology is still descriptive and classificatory, and talk of information flow, and even of goal, often passes for explanation. Still, all biologists pay at least lip service to the synthetic theory of evolution, which is a mechanismic theory, and acknowledge that biological research cannot go far without physics and chemistry in the search for the mechanisms of processes. An increasing number of biologists are working on mathematical models of biological processes on the various biolevels, from the cellular to the ecosystemic. All biologists realize that, in order to know how a bird flies, it suffices to film it in flight; but that, if we wish to understand how it manages to fly, we must create and test models of the mechanics and physiology of bird flight. Likewise psychologists are beginning to learn that behavior, far from explaining, is the thing to be explained: that there are no kinematical explanations but only kinematical descriptions. In particular some of them are learning that only physiology and sociology can explain human behavior. At the same time historians are realizing that chronicling does not suffice: that we can understand historical processes with the help of disciplines other than history, in particular psychology, demography, sociology, economics, and politology. Moreover they have learned that consideration of these disciplines elicits the search for new data as well as the framing of descriptions that would otherwise not be forthcoming.

In sum, in all epistemic fields the tendency has been to advance from description to explanation, from black to grey to translucid boxes, from the ledger to the clockwork. This is no accident but an indication that, although we can understand in various ways, only a mechanismic (or translucid box) theory can give us a satisfactory understanding of things, for it alone can tell us what makes things tick. Not that the above mentioned trend is inevitable: like every other trend it might be reversed. As a matter of fact the current computer fad is already distorting that trend: the lust for computation and simulation dries up the passion for explanation. To be sure we need both computation and simulation, but neither is enough, for neither brings insight. (In particular, the numerical solutions to equations are unwieldly and hard if not impossible to interpret.)

1.2. *Subsuming and Explaining*

Let us analyze three epistemic operations that, though distinct, are often confused: description, subsumption, and explanation. From a logical viewpoint a *description* is an ordered set of factual statements. For example, "The car came at high speed, crashed against the fire hydrant, and stopped". (If the data came in reverse order, the description had to be conjectured.) Given that description we may understand some of the related facts, such as the final state of the driver, the car, and the fire hydrant—but this only because we use tacitly certain generalizations about the effects of impact. But the description does not explain why the driver crashed in the first place: was he drunk, was he sick, did the car have a mechanical defect, was the pavement slippery? Only further data and hypotheses can answer the why-questions.

A *subsumption* too is an ordered set of statements, but one such that the last statement follows from the preceding ones. For example, suppose we want to account for the fact that a given population of bacteria has increased eight-fold in the term of one hour. We investigate the matter, counting (or sampling) the population every so often, and come up with a table or a graph population *vs.* time. Suppose we find that the growth pattern is 1, 2, 4, 8, 16, 32, 64, etc., at 20 minute intervals. This accounts for the fact to be understood, since we recognize it as the 4th member of the sequence. In other words the fact has been *subsumed* under a general pattern, namely: The given population of bacteria doubles every 20 minutes. In other words, we have turned the fact to be explained into a particular case of a law statement.

The inference has been this:

For every x and every t, if x is a population of bacteria of species A in condition B, then the population of x at $(t + 20)$ min is twice the population of x at t. [In general, $N_t = N_0 2^t$, where N_0 is the initial population and t is the number of 20 min periods.]

b at $t = 0$ is a population of N_0 bacteria of species A in condition B.

b at $t = 60$ min is a population of $N_3 = 8N_0$ bacteria of species A in condition B.

Generalizing, the basic subsumption schema is this:

Pattern	*For all x, if Fx then Gx*
Circumstance	*Fb*
Given fact	*Gb*

↓*Deduction*

Sometimes the pattern occurring in a subsumption is merely a systematic or classificatory statement. For example, "Johnny does X because he is (in the class of children who are) 7 years old" is typical of Piagetian psychology. It accounts for a property of an individual in terms of its belonging to a certain kind. And "Helium has a very low reactivity because it is a noble element" accounts for a property shared by all the members of a species by including the latter in a certain genus. In both cases the inference patterns are strictly deductive:

Explanation of individual property in terms of belonging to a species

Premise 1: Definition of species S by the characteristic properties $A, B, \ldots, X, \ldots$ of its members.

Premise 2: b belongs to S.

Conclusion: b possesses property X.

Explanation of species property in terms of inclusion in a genus

Premise 1: Definition of genus G by the characteristic properties $M, N, \ldots, Y, \ldots$ of its members.

Premise 2: S is included in G

Conclusion: Every member of S possesses property Y.

We understand the given fact as a particular case of a general pattern. Still, we do not understand the pattern itself, and therefore the fact to be accounted for remains half understood. We are told how things are, not why they should be the way they are. From a logical point of view bacteria might grow in many different ways or in none. But microbiological research reveals that bacterial colonies happen to grow by division. In our example, every bacterium of species A divides at the end of 20 min; in turn, every daughter bacterium divides after 20 min, so that the original bacterium has been replaced by four, and so on. This then is the (main) growth mechanism in the case of populations of bacteria (and other unicellular organisms): cellular division. (To be sure it is a nonmechanical mechanism.) We have now attained an *explanation* proper, or mechanismic account.

The logical reconstruction of the explanation process is this. From the observation and timing of cell division we hypothesize (and check) the law of growth, namely $N_t = N_0 2^t$. (We can obtain this hypothesis by a rigorous application of the mathematical principle of complete induction. But of course this is no proof that bacteria do grow that way. Only experiment can tell whether the hypothesis is true. In fact we know that it fails for high population densities, i.e. where there is strong competition for resources.) The rest follows like in the case of subsumption.

The difference between subsumption and explanation is not logical: both are cases of deduction from regularities and circumstances, in particular law statements and data. The difference is another one. Subsumption answers only how-questions, explanation how-or-why-questions, such as 'How does this machine work?' and 'Why did this machine go out of order? Both answers or accounts are given in terms of patterns, such as trends, empirical generalizations, or law statements: the given fact to be accounted for is shown to be a particular case of such pattern. But whereas in the case of subsumption the pattern itself remains unaccounted for, in the case of explanation it is a mechanismic hypothesis or theory.

The basic logical pattern of subsumption was

(1) *For all x, if Fx then Gx, Fb* $\vdash$ *Gb.*

The corresponding pattern of explanation is this one:

(2) *For all x, if Fx then Mx, For all x, if Mx, then Gx, Fb* $\vdash$ *Gb,*

where 'M' symbolizes some mechanism such as cell division. (For example, the first premise could be "For every x, if x is a bacterium, then x reproduces by cell division", and the second "If x reproduces by cell division, then the offspring of x grows geometrically".) Now, the two general premises of (2) jointly entail the general premise of (1). Phenomenalists, conventionalists and simplicists would say that this shows that 'M' is dispensable or replaceable. A realist concludes instead that *explanation subsumes subsumption,* logically, epistemologically and ontologically. Logically because, given an explanation, we can detach the corresponding subsumption. Epistemologically because explanation presupposes more knowledge than subsumption. Ontologically because explanation goes deeper into the matter than subsumption, by pointing to some (conjectured or established) mechanism that may be hidden to the senses. Let us dwell for a moment on the epistemological and ontological aspects of explanation that we have just noted.

That explanation involves subsumption is suggested, though not proved, by the following examples. (*a*) Knowing that the mechanism of radio communication is electromagnetic field propagation, we know how it can be produced, how it proceeds, how long it takes—hence when the signal is to be expected—how it may be detected, amplified, etc. (*b*) Knowing that the mechanism of heredity is summarized in the vicissitudes of the parental DNA molecules, we know in principle how hereditary traits appear and fail to appear, and how they could be modified. (*c*) If we knew the neurophysiological mechanism of problem solving we would be able to solve a whole cluster of problems, such as: What particular neural systems work out the problem?, How does the problem solving process proceed, and When do the various stages in the process occur? In sum, knowing the most involves knowing the least. Hence descriptivism—the epistemology that extols description and decries explanation—hampers the advancement of knowledge.

The relations between explanation and ontology are as follows. Firstly, the quest for explanation is a quest for knowing deeper and deeper levels of reality. (For the concept of a level see Vol. 4, Ch. 1.) Secondly, the very choice of a type of explanation presupposes some ontology or other. Thus a theological world view suggests explaining natural and social facts ultimately in supernaturalistic terms, or even holding that they are unexplainable. A physicalist (or vulgar materialist) ontology suggests looking for explanations in physical terms even in the case of biological, psychological, and social facts. On the other hand, emergentist materialism (Bunge, 1981a) acknowledges the emergence of qualitative novelty and it encourages its explanation. (Holism denies that such explanation is possible, and reductionism denies emergence altogether.) Thirdly, to explain a fact is to show how it comes about, not to eliminate it. Thus to explain an illusion is to disclose its neurophysiological mechanism, not to show that it did not occur to begin with. (On the other hand Coren and Girgus (1978, p. 23) hold that "Ultimately, when we know exactly how the visual system works, visual illusions should no longer exist".) In sum explanation, unlike definition, does not eliminate. And inexplicability (or mysteriousness) does not define. For example, *pace* numerous philosophers, "emergence" is not definable as unexplained novelty, and "illusion" is not definable as unreality.

An explanation is an epistemic process involving three components: (*a*) an explainer or animal doing the explaining; (*b*) the object(s) of the explanation, i.e. that which is explained, and (*c*) the explanatory premises or *explanans*. All the explainers we know are humans: we do not know whether any other animals explain, and we will never know unless the question is seriously posed. (True, we often say that a hypothesis or a theory explains such and such facts. But this is by way of metaphor: we mean that a knowing subject explains facts with the help of a theory and data.) The objects of explanation are facts. And the explanatory premises are the hypotheses and data involved in the explanation.

What can be objects of explanation? A sybilline answer is: Anything problematic. But of course what is problematic depends on the general framework as well as on the state of knowledge. Thus a philosopher obsessed with nothingness finds it natural to ask why there is something rather than nothing, whereas a scientist takes the universe for granted and asks instead why things change the way they do, as well as why certain things seem not to change at all. Neither believers nor atheists ask why there is a deity: the former because they take it for granted, and the latter because they deny its existence. The evolutionary biologist asks why the known species have the properties they have, whereas the creationist has no such problem. The nonconformist asks why we should keep certain institutions which the conformist does not question. In sum, each world view encourages the search for certain explanations while discouraging others.

Aside from considerations about world views, the objects of explanation can be things, properties or states of things, or events (changes in things). A (concrete) *thing* is explained if a possible process is found or conjectured that ends up in the emergence of the thing—e.g. the formation of a molecule or of a social system. (Accordingly the stable elementary particles, such as the electrons, are not explainable except when they result from processes such as nuclear disintegration or high energy collision.) Likewise a *property* is explained in terms of some process of qualitative change—e.g. the emergence of the questioning ability during human development. (Accordingly the basic properties of matter, such as that of possessing energy, are not explainable.) And an *event* is explained in terms of forces or of other events, e.g. collision or random shuffling. (Accordingly the uniform motion of a body or of a photon in a vacuum are not explainable: only their origin and their eventual deviation from uniform motion are explainable.)

(Note that we have not included statements nor, a fortiori, sets of statements, among the objects that can be explained. To be sure a statement, unless it is an axiom or a definition, may be deduced from other statements. In particular law statements can often be deduced from higher level law statements. But we had agreed to restrict the meaning of "explanation" to subsumption *cum* mechanism. To be sure, sometimes we deduce law statements from higher level formulas, but not every such deduction amounts to an explanation of the corresponding pattern. For example, every equation of motion and every field equation is deducible from some variational principle, but such deduction does not explain motion or field propagation. On the other hand the deduction of a "deterministic" equation of motion or some other pattern from probabilistic laws does amount to an explanation of the "deterministic" pattern. Likewise we explain behavioral regularities in neurophysiological terms. In sum, deduction may or may not correspond to explanation.)

Obviously not everything explainable is worth explaining. The value of an explanation depends on the world view, the state of knowledge, or practical considerations. For example, sometimes we need to explain regularity—e.g. why Johnny has above-average school attendance; at other times anomaly—e.g. why Johnny, who so rarely misses school, is missing today. When we know or suspect that a given process is causal, the observation of accidental departures cries for explanation; but if the process is assumed to be random, then we must explain its regular features. And whereas some students take certain data for granted, others try to explain them. Different people itch in different places at different times.

We have drawn a sharp distinction between explanation and subsumption, suggesting that the latter is little more than description. This distinction is not made in the philosophical literature (except for Bunge, 1967b). The reason for this is that most philosophers have focused on the logical form of explanation, disregarding its epistemological and ontological aspects. This neglect has suggested to some philosophers that the standard view on explanation, found in Popper (1959) and Hempel (1965), is not just incomplete but basically mistaken. In particular it has been suggested that explanation is not a hypothetico-deductive operation but consists in drawing analogies or metaphors, or in empathizing with the actors of the psychological or social phenomena to be explained. These criticisms rest on a confusion between the psychological category of understanding and the methodological category of explanation. As noted in Section 1.1 we may indeed understand via metaphor or *Verstehen*, but explanation is different.

What is true is that most of the explanations offered in ordinary life and even in science and technology are incomplete, and so seem not to match the hypthetico-deductive model. In fact they presuppose (assume tacitly) a number of hypotheses. The most famous of all such incomplete (or enthymematic) arguments is Descartes' "I think, hence I exist (I am alive)". The tacit premise here is "All thinking beings are alive", which, together with the explicit premise "I think", yields the conclusion "I am alive". So although it looks like a proposition it is actually an abbreviated argument.

Incomplete explanations, or *explanations in principle*, abound in all fields of inquiry. Thus the technologist is often content with pointing to "the principle" of his design: he invokes a law statement, or a theory, to explain in outline why his artifact should work. The economist who succeeds in explaining an economic trend does so in principle, not in detail, because his equations (even supposing they were accurate) contain far too many unknown parameters (Hayek, 1955). Likewise the evolutionary biologist can explain in principle the evolution of some biopopulations with the help of the general principles of the synthetic theory of evolution. A full explanation of the evolution of any particular biopopulation would require a precise mathematical model involving mutation probabilities, viabilities, and environmental factors such as the competing populations. However, even a partial explanation is, or can be cast in the form of, a deductive argument: *Pattern and circumstance, ergo explanandum* (fact to be explained).

We conjecture that every explanation, if properly analyzed, will prove to be *incomplete* in failing to explain, with total accuracy, every feature of the object of explanation. This is but one aspect of the incompleteness of all our factual knowledge. Irrationalists should derive no comfort from this postulate, for we also conjecture that every explanation is *perfectible* with the help of further research. Such research generates *explanation chains* of the form *A because B because C*... *M because N*. For example, a girl goes to school because she wants to learn. She wants to learn because she needs a good job. And she needs a good job because her parents cannot support her. An explanation chain is a partially ordered set of explanatory statements of the form "*X* because *Y*". And, as we saw a moment ago, every statement of this type is an enthymeme analyzable into an argument.

Is there a limit to every explanation chain? That is, are there ultimate facts such as Aristotle's Prime Mover? We have no evidence for such ultimates. All we know is that no explanation satisfies forever. For example, Newton explained the planetary orbits in terms of inertia and gravitation. His adversaries, in particular Leibniz and the followers of Descartes, were dissatisfied with Newton's hypothesis of gravitational interaction at a distance: they wished to explain gravitation itself. This explanation came much later with Einstein's field theory of gravity.

Still, there are dogmatists who proclaim that whatever we cannot explain satisfactorily now will remain mysterious forever. One such mystery would be the precise conditions of the "big bang", for the evidence needed to investigate the cause of the primal explosion was destroyed in the explosion itself (Jastrow, 1978). Yet it may well happen that such evidence will eventually be discovered, or even that evidence against the "big bang" hypothesis will be produced. Another alleged perennial mystery is that of consciousness (Eccles, 1978). It is more fruitful to assume instead that this problem had been approached in the wrong ways, namely via theology, philosophy, and mentalistic psychology rather than neuropsychologically. Even if no satisfactory solution to the present mysteries were forthcoming, we would learn more by investigating them than by declaring them to be insoluble mysteries. Research pays more than dogma.

Explanation dissolves mystery but need not remove marvel. The widespread fear that scientific explanation may suppress marvel and our capacity to enjoy nature, art, and human relations, is groundless. On the contrary, (*a*) searching for explanation and finding it is in itself an enjoyable experience; (*b*) by expanding and deepening knowledge, scientific research allows us to marvel at things and processes that were unsuspected before; and (*c*) the practical fruits of science allow millions of people the leisure necessary for enjoying nature, art, and human relations.

Nor does explanation reduce the unfamiliar to the familiar. On the contrary, science explains many familiar phenomena, which we take ordinarily for granted, in esoteric terms. Think of the way meteorology explains the familiar facts of weather, the way chemistry explains the transmogrification of food in the kitchen, the way psychology explains perception, memory and thought, or the way economics attempts to explain inflation. It all sounds, to the layman, very much like the *explicatio obscurum per obscurius* that the modern philosophers attributed to the schoolmen. But at least it is not an *explicatio ignotum per ignotius*: on the contrary, we can now explain the unknown in terms of the known or conjectured. Moreover, obscurity is relative to the subject: in principle anyone can master the technicalities of science. (By the way, the fashionable claim that the use of technicalities by scientists is a manifestation of elitism, which a science for the people should avoid, is sheer obscurantism. There is no deep knowledge without tears. The remedy is not to degrade science but to upgrade education.)

Another widely held belief is that an explanation is nothing but an answer to a why question. Let us test this account with a simple example: Why do people who walk on snow wear snow shoes? Answer: So as not to sink. No doubt, this answer is correct—but it does not explain anything. An explanation requires a more complete answer, such as this one. One sinks in snow the deeper, the harder one presses on it; snow shoes, by distributing the weight over a greater area, decrease the pressure and so also the sinking. In short, although every explanation answers a why (or a how) question, the converse is not true. Moral: Ordinary language is no substitute for conceptual analysis.

We wind up this section with a handful of maxims or rules for explanation.

R1: *Before rushing to explain a fact make sure it is fact not illusion.* (E.g. before explaining spoon bending at a distance establish that it is a fact not a fraud.)

R2: *Explain existents by existents* (established or conjectured) never by nonexistents. (Hence mistrust the explanation of elementary particle events in terms of virtual particles or processes, and of human behavior in terms of disembodied souls or immaterial processes.)

R3: *Explain the observable by the unobservable* (e.g. color changes by chemical reactions) *or the unobservable by the observed* (e.g. societal facts by individual behavior) instead of keeping them separate.

R4: *Mistrust ad hoc explanations*, i.e. explanations with the help of hypotheses that cover only the fact to be explained.

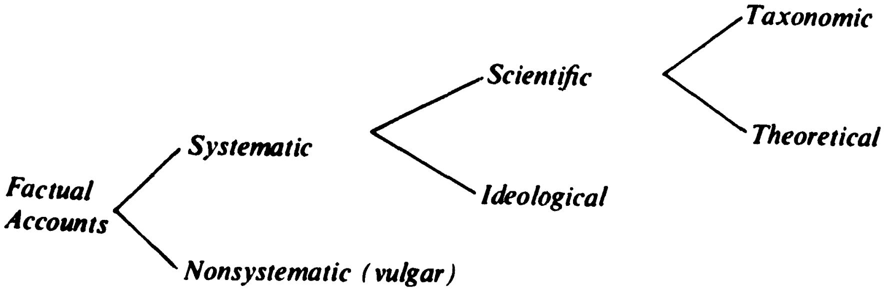
R5: *Mistrust hypotheses and theories that purport to explain everything* (such as psychoanalysis, which professes to explain all human life, and psychoneural dualism, which professes to explain all neural and mental processes).

Only one type of explanation satisfies all of the above rules, namely that performed with the help of scientific theories. Let us then have a look at it.

2. SYSTEMATIC ACCOUNT

2.1. *Vulgar, Ideological, and Scientific Accounts*

Descriptions, subsumptions, explanations, and forecasts will be called collectively *accounts*. Typically, accounts involve hypotheses as well as data; and, whereas some hypotheses are testable and even true, others are just myths. An account can be systematic or nonsystematic. A *systematic account* is one involving some conceptual system: a classification, a theory, or part of either. The system may be scientific or ideological (religious or sociopolitical). If the former, the account will be called *scientific*, if the latter, *ideological*. A nonsystematic account employs only ordinary knowledge (true, half true, or sheer superstition); it may also be called *vulgar*. In short, we have the following classification of accounts of matters of fact:



Examples of each of these categories abound. Thus when explaining a young man's absent-mindedness by his being deeply in love we may be right, but such explanation is nonsystematic: it makes use of a piece of popular wisdom ("People in love tend to act absent-mindedly"), not a conceptual system. The fact that some children resemble some of their long deceased ancestors can be explained either ideologically, by reincarnation, or scientifically, by the genetic hypothesis of recessive genes. Survival after death may be accounted for by the doctrine of the immaterial and immortal soul. In this case there is no rival scientific account, for there can be no evidence for survival after death: to science this is not a fact, hence it does not call for an explanation. What does call for a scientific account is the origin and persistence of the belief.

The typical vulgar account is *ad hoc*, i.e. concocted for the occasion, and *unifactorial*, i.e. it involves a single factor or variable. On the other hand every scientific account is multifactorial, i.e. it involves at least two variables. The reason for such multifactoriality is that every variable is related to at least one other variable, and usually to more than one; so, any change in one of them is bound to affect others. However, because unifactorial accounts are the simplest and have the greatest (though illusory) unifying power, even scientists fall occasionally for them. One instance is the attempt to explain all behavior and all thought in terms of language (Luria, 1961). Another is the cognitivist account of all cognition as rule-directed computation of the same kind as that performed by computers (e.g. Pylyshyn, 1980). A third example of a unifactorial account is the explanation of social change, hence of history, in terms of imitation or innovation, class struggle or political strife, population growth or environmental change, the power of ideas or that of rulers. Seasoned historians know that history is far too complex a process (or bundle of processes) to be explainable by a single factor. They know that social change may start in the economy, or in the culture, or in the polity, and propagate to the other subsystems of society: they know that there is no absolute prime social mover. Likewise biologists explain evolution in a multifactorial fashion: in terms of chance genic variations, geographical isolation, climatic change, ecological catastrophes, natural selection, etc.

Ideological accounts too tend to be unifactorial, but they are all-embracing rather than *ad hoc*: because of this simplicity and unifying power they are as dangerous as they are attractive. Thus political credo X-ism blames all the evils of society on anti-X-ism, and conversely. Astrology attributes every event in human life to the stars. Psychoanalysis blames every psychological trouble on early sexual experiences, whether real or imaginary. Interactionist mind–body dualism (e.g. Popper and Eccles, 1977) accommodates every possible mental event, and so dispenses with all specific theories (of e.g. perception, memory, imagination, and thought) as well as with experiment. (But it explains nothing, for it just states—in an inexact way to boot—that mind and body interact in such a way that they pass one another the buck, without suggesting any interaction mechanism.)

The believer in an ideology has the great advantage, over the vulgar, of being able to offer a single account of every single fact, and of finding it ready made in his system instead of having to search for it. In an open context, on the other hand, every fact can be accounted for in infinitely many ways. Thus if a fact described by a proposition q is observed to happen, then we can deduce that q because p, where p is an arbitrary proposition. (Indeed, the inference "q entails If p then q" is valid in ordinary logic.) Not so in a closed context: here only propositions logically and referentially germane to the given proposition q can be invoked. Moreover in a theory, whether ideological or scientific, there is always at most one explanation for each observation. Therefore, paradoxically, not only the ideologist but also the scientist are spared the agonizing uncertainties of the vulgar: the former because he knows all the answers in advance, the latter because he is confident that he, or someone else, will eventually come up with a satisfactory answer if he investigates the matter. Both are sustained by faith—irrational in the first case, rational in the second.

Scientific accounts are or tend to be systematic and multifactorial, and they avoid the two extremes of the *ad hoc* (or single case) and the all-embracing accounts. Of course there are some extremely general scientific systems, such as the relativistic theory of gravitation and the neo-Darwinian theory of evolution. However, they do not embrace every possible fact in all possible domains, and they do not propose essentially the same account of every fact. (Anything that purports to explain everything actually explains nothing.) Indeed only very specific theories, i.e. models, can account for facts without much further ado. Any account of specifics done with the help of a general scientific theory requires enriching the latter with subsidiary assumptions and data. Let us make a quick study of three kinds of scientific account: description, subsumption, and explanation; prediction will occupy us in Section 4.

A description will be said to be a *systematic scientific description* if it is couched with the help of a scientific system. In particular, a *scientific taxonomic description* is made with the help of some scientific classification, such as one of elementary particles, plants, or economies. *Example*: "Humans evolved from more primitive primates about three million years ago". A *scientific theoretical description* is one made with the help of a scientific theory. It consists in specifying the values of the state variables of a thing as hypothesized by the theory. *Example*: "The electromagnetic wave, of 1 cm long wavelength, hit the two slot screen; the waves emerging from the latter interfered with one another, and were absorbed by a second screen lying 10 m behind the first". Contrary to the empiricist prescription, scientific systems are not built on presystematic descriptions: on the contrary, scientists describe their facts with the help of their conceptual systems.

Caution: Not everything that sounds theoretical is so. Thus over the past few decades a number of scientists have tended to account for facts in the language of the statistical theory of information. They talk about information flows between molecules, the quantity of information locked in a gene, the way the brain processes the incoming information, information networks in society, and so on. Such accounts do not really use the statistical *theory* of information but only its *vocabulary*. In fact, typically they do not identify the sender, the channel and the receiver; they do not assume that the channel is subject to random perturbations; and they do not calculate or measure quantities of information. (As one biologist put it, "it is doubtful whether information theory has offered the experimental biologist anything more than vague insights and beguiling terminology": Johnson, 1970.) Such accounts are therefore *pseudotheoretical.* To be sure they had initially, particularly in genetics, neuroscience, and psychology, some heuristic value. But they have now become obstacles to theorizing in those fields, because they suggest that the theory—namely the statistical information theory—is already there. Besides, even if one were to make use of information theory in those fields one would achieve at most an exact description: information theory is far too general to explain or predict anything. The same holds for other general system theories: the very notion of a general system theory of a particular system, such as a cell, a lake, or a school, is self-contradictory. An extremely general theory is at best a basis for building a specific theory or model. (Recall Ch. 9, Section 1.2.)

Let us now turn to subsumption. As we saw in Section 1.2, a fact (or the proposition describing it) is subsumed under a pattern (or the corresponding proposition) if shown to be a particular case of it. A subsumption will be said to be a *systematic scientific subsumption* if the subsuming construct is a scientific system, such as the periodic table of the elements or the input–output model of an economic system. In particular, a *scientific taxonomic subsumption* is one where the subsuming construct is a scientific classification. *Example*: "Copper, silver and gold are simple metals". A *scientific theoretical subsumption* is a subsumption under a scientific theory, i.e. such that the explanatory premises are theoretical formulas and data. For example, every instance of light refraction falls under Snell's law. In particular this law, together with the values of the angle of incidence and the refractive indices of the two adjoining media, entails the value of the angle refraction and thus accounts for it. But, because the Snell law is phenomenological (or black box), nothing gets properly explained in this way: the account looks very much like a description, except that it involves a deduction.

(The logic of subsumption boils down to the following. Let $\mathcal{T}$ be a scientific theory in a context $\mathcal{C} = \langle P, Q, R \rangle$, i.e. such that (*a*) the set of formulas of $\mathcal{T}$ is a subset of P, and (*b*) some formulas in $\mathcal{T}$ refer to entities in R. Further, let p and e be propositions in P but not in $\mathcal{T}$, and such that (*a*) p describes a fact or class of facts f involving entities in R, and (*b*) e is a bit of empirical evidence relevant to p (and therefore referring to R's or to things, such as measuring instruments, that are somehow connected with R's, as is the case with telescopes aiming at stars). We stipulate that $\mathcal{T}$ *subsumes* f iff there is in P, but not in $\mathcal{T}$, a (possibly empty) set S of subsidiary assumptions such that p follows from $\mathcal{T}$, S and e.)

The fact being subsumed may be some thing being in a given state or undergoing a given change. If a whole class of facts is being subsumed, then the proposition that is being derived with the help of the accounting theory is a hypothesis not included in exactly the same form among the general principles of the theory. For example, the various laws of geometrical optics—rectilinear propagation in a homogeneous medium, mutual independence of the components of a light beam, reflection and refraction of rays—are subsumed under wave optics. In geometrical optics the refraction index is an irreducible phenomenological parameter; in wave optics it is a kinematical parameter, namely the ratio of the velocities of propagation in the two media. In electromagnetic optics the refractive index is the former as well as a function of the molar electric and magnetic properties of matter; the index is explained properly only by the theory of the atomic structure of matter.

We construe explanation as a particular case of subsumption, namely as *subsumption under a mechanismic* (or translucid box) *theory*. Indeed we stipulate that a fact or class of facts F is *explained* by a theory $\mathcal{T}$ iff (a) f is subsumed under $\mathcal{T}$ and (b) $\mathcal{T}$ is mechanismic (instead of phenomenological). This definition covers the explanation of factual items of all kinds: properties (hence states) of things, events (changes) in things, and patterns of things, such as trends and even laws. Thus we say that the theory of electron dispersion explains the refractive index of quartz; that the electromagnetic theory of light explains the refraction process; that quantum mechanics explains the formation and breakdown of molecules; and that economic theory is supposed to (though it is doubtful that it does) explain the inflation characteristic of the 1970s.

Note the following points. First, every explanation is a subsumption but the converse is false. Second, explanation is richer than subsumption, which is only a special case of deduction: explanation tells us not only what happens but also how it happens. Therefore not all theoretical models are explanatory. In particular, a mere mathematical representation of a set of facts, even if true, does not explain anything. Third, whereas in physics, chemistry, and biology, explanation is nomological (i.e. involves law statements), from psychology upwards rules and evaluations may be involved as well. Thus we may explain somebody's action A by saying that A

causes B (*law*), the individual knows this and values B (*valuation*), and he has adopted the *rule* or maxim that he will always endeavor to attain what he values. Fourth, theories about matters of fact contain law statements and they may also contain valuations and rules, but they do not contain trends or changes in trends. Far from explaining anything, trends and their changes are objects to be explained in terms of laws, valuations, rules, or circumstances. For example, the steady increase in the international price of crude oil is explainable by the decision of the OPEC countries to stop financing the affluence of the OECD countries. And the growth of the tertiary sector of the economy (i.e. the services) is explainable by the increasing importance of technology in agriculture and industry, as well as by the passing of welfare legislation.

I submit that the above explicates what scientists usually mean by 'explanation', namely an account in terms of some mechanism or other, be it causal or stochastic, physical, biological, social, or other. However, once in a while pseudoexplanations appear in the scientific literature. For example, a biologist may lapse into teleology when forgetting that no modern biological theory (in particular no mathematical model) contains the concept of goal or final cause. Or he may be tempted to explain morphogenesis in terms of "morphogenetic fields", ignoring that, in order to speak literally about a field, there must be a field theory with definite and testable field equations. Likewise the computer scientist and the cognitivist (or functionalist) psychologist may feel that they explain the mind when translating mentalistic psychological descriptions into the language of computer science—e.g. when they translate 'x recalled y' into 'x retrieved y from his memory bank'. Translation into one's professional jargon produces familiarity not explanation.

We have called 'theoretical subsumption' what most philosophers call 'explanation'. The classical account of that operation, though one different from ours, is that of Hempel and Oppenheim (1948), usually called 'the deductive-nomological (or D-N) model of explanation'. It has been criticized in various ways, some correct, others incorrect. Ordinary language philosophers have complained that in ordinary language one often calls 'explanation' statements of the form "q because p", which are not logical arguments and need not mention any law statements, so that they are true of false. This criticism, like that of the use of logic in philosophical analysis generally, misses the point: philosophers of science are interested in a technical concept of explanation, which may be absent from ordinary knowledge considerations but happens to occur in scientific contexts.

A relevant criticism is that the Popper–Hempel–Oppenheim relation of explainability holds between almost any general hypothesis and almost any singular sentence (Eberle, Kaplan and Montague, 1961). Consequently there is no great merit in subsumption—not more than in curve fitting, a close relative. Weightier criticisms than this can be raised against the D-N "model" of explanation. One is that it usually requires the explaining premises to be true. This requirement is unrealistic, for in science and technology total truth is hard to come by: our hypotheses and theories are at best partially true, and sometimes our data too are far off the mark—to the point that we may seek to subsume "facts" that, under closer examination, prove to be nonexistent. (Strangely enough most of the repairs proposed to the D-N account of subsumption keep the condition that the explaining premises be true.) A second defect of the D-N "model" is that it is about an open-context operation: it does not require the explanatory premises to belong to theories (hypothetico-deductive systems). But, as we have seen, in an open context almost anything may be accounted for in a number of alternative ways: it is only relative to a definite conceptual system that one can speak of *the* subsumption, or *the* explanation, of a fact. Thirdly, the D-N "model" does not mention the subsidiary assumptions that occur in almost every systematic account of facts but are not part of the theory itself—e.g. the composition and structure of the system to be accounted for. (This oversight may be due to the well-known fact that most philosophers of science think of classical particle mechanics when they write about theories.)

But the most damaging criticism that can be made of the D-N "model" of explanation is one that is seldom voiced, namely that it is concerned with theoretical subsumption, not with genuine explanation: it ignores the mechanisms of things. In this regard—not with regard to the logical apparatus—it is less satisfactory than the classical account of explanation, from Aristotle to Meyerson (1921). According to this alternative view we explain events by their causes, and if we do not know the causes then we may describe or subsume but not explain. As Whewell (1847, P. I, p. 652) put it, "No sound theory without Aetiology". (For an exposition and criticism see Bunge (1959a).) This doctrine became obsolete the moment probabilistic theories began to invade science and technology. We submit that a genuine explanation is one that invokes some mechanism or other, whether causal or stochastic (probabilistic). We shall devote the next section to these two modes of explanation.

We close this section with some thoughts on the power of theoretical accounts in general, be they descriptions, subsumptions, explanations, or forecasts. Two rival accounts of the same fact, or class of facts, may not be equivalent: one of them may be better than the other. Because rationalists are supposed to justify their valuations, let us explicate the notion of preference involved in the statement schema "Account X of fact f is preferable to account Y of f". We shall say that a theoretical account is the better, the wider its coverage of facts, the more accurate, and the deeper.

(More precisely, let A_1 and A_2 be two accounts the reference classes of which have a nonempty intersection R. We stipulate that (*a*) A_1 has a greater *range* than A_2 iff the hypotheses of A_1 are more general (less *ad hoc*) than those of A_2; (*b*) A_1 has a greater *accuracy* than A_2 iff A_1 involves truer premises than A_2 and, in the case of prediction, A_1 scores better (is better confirmed) than A_2; and (*c*) A_1 is *deeper* than A_2 iff A_1 involves more levels of organization than A_2. These three aspects are mutually unrelated: depth is compatible with either generality or specificity, as well as with accuracy or inaccuracy. The *best* of two roughly equally accurate and equally general accounts is the deeper, and the best of two equally deep accounts is the more accurate and general one.)

2.2. *Basic Explanation Types*

We have stipulated that, unlike a description, a subsumption, or a prediction, an explanation invokes some mechanism or other (Section 2.1). We may distinguish five basic or pure types of mechanism or mode of becoming: causal, stochastic, synergic or cooperative, dialectical or conflictive, and teleological (Bunge, 1981a). These five pure types may combine by twos or more to form a total of 26 mixed types. So, there are at least 31 types of mechanism, or modes of becoming, hence 31 types of theories and models, and therefore 31 types of explanation. This number would decrease if it were proved that some of the modes of becoming which we have regarded as basic are reducible to others; but future research might also increase the number of basic types of mechanism.

It seems safe to assume that the pure types occur in reality only approximately: that only mixed types of becoming occur in nature and society (Bunge, 1959a). For example, the processes accounted for by the quantum theories are causal as well as stochastic, at least wherever forces (or potentials) are involved (Ch. 8, Section 3.1); every realistic theory of systems of any kind must take into account cooperation (in system formation and maintenance) as well as conflict (among system components or between them and environmental items); evolutionary processes involve causation, chance, cooperation, and competition; and human social processes involve all five basic types. However, most theories and theoretical models in science and technology involve only one or two basic types of mechanism. This restriction is sometimes due to the fact that there is often only one, or a couple, of overriding mechanisms; at other times it is merely the result of ignorance or even prejudice. Let us make a quick review of such models and the corresponding explanations.

Causal mechanisms, and the accompanying explanations, are probably the most familiar. There is causation if some properties of a thing change under the influence of external events. The former event (or process) is called 'effect', and the latter 'cause'. (These concepts can be elucidated in terms of the notion of history of a thing: the total effect is the distortion in its history due to external influences. See Bunge, 1982.) The best known example of a causal theory is automata theory. Indeed a deterministic automaton stays in a given state unless it accepts an input, and when it does it jumps to another state. No input, no change of state and therefore no output. (Formally: the next state function M, that maps the set of ordered pairs $\langle$state, input$\rangle$ into the set of states, satisfies the condition $M(s,0)=s$, where s is an arbitrary state and 0 the null input.) Only such changes of state are in need of explanation, and they are explained in terms of inputs and previous internal states. Similarly in mechanics all causes are forces or constraints, and the corresponding effects are deviations from uniform and straight line motion, or changes in inner stresses. Only such deviations, not motion itself, are to be explained. (However, mechanics is not a purely causal theory, if only because, unlike automata theory, it involves self-movement in the guise of inertia: see Bunge, 1959a.) Similar causal or partially causal mechanisms, some mechanical and others not, are hypothesized in all fields of factual research.

All events involving large scale systems have joint multiple causes. There are two ways of disentangling such causal factors, i.e. of imputing the observed or conjectured effect to the various causes at play. One is to "freeze" experimentally all factors but one, by using such procedures as insulation and compartmentalization—which are not always practicable. The other method is to perform a theoretical analysis. For example, the radio waves emitted by radio stars lying behind the sun are deflected by the latter. This deflection is due partly to refraction by the solar atmosphere (corona) and partly to the solar gravitational field. The way to separate these effects is to use the knowledge that refraction depends on wave length, so that the total deflection equals the sum of the gravitational deflection and a wavelength-dependent deviation. By measuring the total deflection for different wavelengths one can identify the gravitational part of the total effect.

The philosophical literature on causality is rather meagre both quantitatively and qualitatively. (For a review see Bunge, 1982.) The two most popular views on causation are the logical and the probabilistic ones, neither of which is adequate. According to the former the causal relation is faithfully represented by a conditional proposition relating the cause to its effect. On this view a cause is just a sufficient (or perhaps necessary and sufficient) condition for its effect. This view is wrong for two reasons: one is that causes are events not reasons, not even sufficient reasons. (See Bunge, 1959a.) Another is that, given any event, one may invent any number of "causes" capable of "explaining" it without violating logic, though in most cases science will be violated. Indeed, a proposition *e* describing the given effect entails $\ulcorner e$ or not-$c\urcorner$, where *c* is an arbitrary proposition describing any old "cause"; but $\ulcorner e$ or not-$c\urcorner$ is equivalent to $\ulcorner c \Rightarrow e\urcorner$. So, causal "explanations" out of context are cheap, hence we should not care for them. What science does cherish is systematic or theoretical explanations, whether or not they involve causal laws. In other words, only if $\ulcorner c \Rightarrow e\urcorner$ is a law statement (hence one included in a confirmed theory), does *c* explain *e*.

The probabilistic view of causation regards the latter as a particular case of a probabilistic relation. (See Suppes, 1970.) The heart of this view is the definition: "*c is the cause of e* $=_{df}$ The probability of *e* given *c* equals 1", or some weaker version, such as "*c is a cause of e* $=_{df}$ The probability of *e* given *c* is greater than the absolute probability of *e*". Two trivial counterexamples are enough to ruin this view. The probability of my dying, given that I was born, equals unity—yet my birth could not possibly be regarded as the cause of my death. Second counterexample: The probability of rain, given a sudden drop in the barometer reading, is greater than the probability of rain *tout court*—yet the barometer reading is just an indicator of rain, not its cause. (For further criticisms see Bunge, 1973e.) In sum, causation is not definable in terms of probability. Nor is the converse reduction of probability to causation possible. What is possible and useful is to relate the two concepts.

We use the concepts of causation and probability on a par in science and technology. Thus geneticists speak of the probability that radiation of a given kind will cause a certain mutation. In the quantum theories we compute not only probabilities of spontaneous (uncaused) events, such as the spontaneous radiative decay of atoms. We also compute the probabilities of caused events, e.g. that a given external perturbation will cause the atom to emit a photon of a given frequency. The concepts of cause and effect are retained in the quantum theories but they are related in a probabilistic, not in a causal way. (Recall Ch. 8, Section 3.1.) In general, in the context of stochastic theories, instead of asking "Does *c* cause *e*?", we ask "What is the probability that *c* causes *e*?" Correspondingly, pointing to the cause is necessary but not sufficient to explain: we must add the probability of its efficiency. In other words, we need two concepts instead of one: those of cause and of efficiency of a cause, which is in turn reduced to that of probability (that the cause will produce its effect).

A probabilistic subsumption is one from premises at least one of which is a probabilistic law (not just a statistical generalization). Thus given a velocity distribution we can infer the average velocity and the spread around that average. (Actually we can infer infinitely many statistical parameters, namely all the moments of the distribution.) This will not explain anything unless it is explicitly stated that the distribution is one of entities moving randomly. The least a probabilistic subsumption does is to show events of a certain kind to be possible, i.e. to have a nonvanishing probability. Such account is of course far poorer, and therefore less satisfying, than the proposal of a definite random mechanism, i.e. a stochastic explanation proper.

Consider the following examples of random mechanisms. The inner face of a window pane becomes fogged in cold days. This process (or the corresponding proposition) can be subsumed under a rate equation representing the increase in thickness of the layer of water vapor on the pane. But this does not explain anything. The explanation is that the water molecules and droplets move randomly about the room and arrive independently from one another to the cold pane, to which they transfer their kinetic energy. This uniformity, and the huge numbers, explain the uniformity of the fogging. Second example (P. and T. Ehrenfest): A dog ridden with fleas meets a clean dog. After a while each dog has approximately the same number of fleas: the two-dog system has attained the equilibrium state. The explanation is that the fleas jump more or less at random from one dog to the other until, in the end, about equal numbers of fleas jump in both directions. Third case: the nearly equal distribution of sexes in humans and other animals is explained by the chromosome composition and the random shuffling, at fertilization, of the genes composing them.

A last example, which is also one of exactification. We all know that, when a message is transmitted along a noisy channel, such as a long distance telephone line, the receiver is likely to fail to understand some words. As the message is repeated, its intelligibility increases; intuitively one expects its intelligibility to be proportional to the number of repetitions. The stochastic model that explains this process boils down to the following. Suppose a given message is composed of N words, and pretend that all of them have the same probability p of being understood by the receiver. The expected number of words that are identified correctly on a first presentation is Np; consequently the number of failures is $(1-p)N$. If these wrongly understood words are repeated, we may expect to understand, on the average, $p(1-p)N$ of them. Therefore after two presentations the total expected number of correctly understood words will be $pN+p(1-p)N$; hence the expected ratio of understood words to total number of words is $p+p(1-p)=1-(1-p)^2$. By iteration we find that, after the nth presentation of the message, the cumulative fraction of correctly understood words, or total intelligibility, is $I_n=1-(1-p)^n$. As suspected, this number is close to np. (An obvious generalization is to unequally probable words. And a deeper model would treat p not as an ultimate but as a function of certain properties of both channel and receiver.)

In all cases of probabilistic explanation what is explained is some statistical regularity, such as an average or a correlation (e.g. a trend). Thus we explain the fact that nearly all individuals in a crowded sidewalk, and nearly all cars on a very busy road, have roughly the same speed, by pointing out that a faster pedestrian (or car) would be stopped, and a slower one would be pushed. (The theoretical problem is: Given the average $\bar{F}$ of a certain magnitude F, conjecture some probability distribution p of F with mean $\bar{F}$. This is an inverse problem and therefore has no unique solution. The direct problem, which has a unique solution, is: Given p compute $\bar{F}$.) What holds for averages holds, *mutatis mutandis*, for spreads and other statistical parameters. In the case of statistical correlations we may try either of the following tactics. One is to assume an underlying causal relation, e.g. via a third factor upon which the two given variables depend. If this move fails one may try a stochastic model including one or more probability distributions allowing one to compute the observed correlations. Comparison with the data will either confirm the model or call for its modification—or for a new, more extended or more precise run of observations. (The correlation may have been accidental.)

Unlike causation and design, chance is often regarded as unintelligible, hence stochastic models as merely temporary devices incapable of explaining anything. Correspondingly only causal (and possibly also teleological) explanations are regarded as genuine. (Thus Salmon (1977, p. 162): "To give a scientific explanation is to show how events and statistical regularities fit into the causal network of the world".) But after more than fifty years of highly successful quantum theories—and thirty of failure to produce any worthy rivals containing only "hidden variables"—it is time to take chance as a basic mode of becoming. In fact these theories, in the realist interpretation, assume that every single entity, be it a particle or a field quantum, possesses ("obeys") objective probabilistic laws, such as position and momentum distributions (Bunge, 1967c). This type of chance is radically different from the accidental crossing of two or more initially independent causal lines—the kind of chance known to the Stoic Chrysippus and inherent in classical statistical mechanics. In the quantum theories we understand events in terms of an interplay of causation and chance.

The philosophical literature on probabilistic explanation is disappointing: it simply does not concern this type of explanation. Thus Hempel (1965, Ch. 12), in a pioneering study on the matter, made an assertion that has been repeated uncritically by a number of other philosophers. He stated that what he called "the inductive statistical" model of explanation differs from the "deductive-nomological" one in two respects. First, the laws are "statistical rather than universal"; second, the explanandum follows from the explanans with high probability rather than deductively. These two propositions are false.

The first proposition involves a confusion between a probability *law*, such as the Maxwell–Boltzmann law or the Schrödinger equation, and a mere statistical generalization such as "f percent of the population of X are illiterate". The stochastic laws occurring in science and technology contain universal quantifiers ranging over their referents, instants of time, etc. (For example, "For all systems of particles in equilibrium, the probability density is proportional to exp $(-E/kT)$".) As for the second assertion we are discussing, all the inferences made in stochastic theories or models are strictly deductive, for they consist either in proofs or in computations. Thus averages are computed from distributions and the definition of an average. In addition to ignoring the probabilistic laws of science and technology, Hempel has confused probability theory (a chapter of pure mathematics), mathematical statistics (an application of the former), and theories of random processes (distributed among the various chapters of science and technology). I have singled out Hempel's study not because it is the worst but because it is the clearest and it has inspired an uncounted number of equally wrong-headed treatments of an important problem.

Now a few words about synergy and dialectics, of which more elsewhere (Bunge, 1981a). Every process of accretion, clumping, self-assembly, or merger, be it of atoms, molecules, organisms, or social systems, is one of cooperation—mostly unwitting to be sure. Thus we explain the formation of biomolecules light and heavy as a self-assembly process, and in turn explain this process in terms of physical forces and chemical bonds. Likewise we explain the breakdown of systems on all levels in terms of conflicts among their components, or among these and environmental things. In turn, a conflict may be explained or analyzed as some dissociation process such as receding motion or competition for a resource.

Both modes of becoming or change mechanisms, cooperation and competition, coexist and intertwine on all levels (Bunge, 1976). Consider two simultaneous chemical reactions of the forms "$A + B \rightarrow C$" and "$B + D \rightarrow E$". In each of these processes the reactants may be said to cooperate to form the reaction product; but A competes with D for B. Hence this is a case of mixed mechanism. (The entire chemical kinetics of systems of reactions is concerned with cooperation-competition processes.) Certainly, saying that C is the outcome of the cooperation (or competition) of A with B explains the emergence of C, but in a superficial manner. The cooperation (or competition) itself can in turn be explained, at least in principle, in terms of forces, constraints, interests, or what have you.

The last type of mechanism in our list is goal-seeking (or teleological) behavior. Thus when dealing with higher vertebrates we often account for animal A's doing B by saying that B serves to attain C, which we assume tacitly to be of value to A. The explicit argument is: "A expects that doing B will bring forth or favor the occurrence of C. Now, A values C. Hence A does B." This is not a deductive argument. It is not even intuitively correct, for A might abstain from doing B, although it values the outcome C, because it is prevented from acting, or because it lacks the means, the energy, or the courage to act. To be sure, we may complete the account by adding that A is free and capable of and willing to do B to attain C. But even so completed the conclusion that A does B does not follow deductively from the premises. So, teleological accounts are not arguments proper, hence they do not qualify as subsumptions, let alone as explanations. In order to transform them into such we need something like a calculus of means and ends yielding the desired conclusions from premises consisting in law statements, rules, and value judgments. (See Bunge, 1977e.) As long as no satisfactory calculus of this kind is available we should avoid teleological accounts, or at least abstain from attributing to them any explanatory power.

3. UNIFICATION

3.1. *Reduction*

Genic variation has been explained as change in the composition or in the structure of genetic material, i.e. DNA molecules. This new account of a previously known process has enabled biologists to effect the reduction of a part of genetics to molecular biology (or biochemistry). In this case the key to understanding has been reduction; in others it is integration. Thus bioevolution is not understandable on the molecular level alone, but calls for a merger of genetics (both molecular and populational), anatomy, physiology, ecology, biogeography, palaeontology and systematics—actually all the branches of biology. In this section we shall study reduction, a much misunderstood operation; we shall leave integration, a generally ignored operation, for the next section.

We can reduce concepts, propositions, and theories. In some cases reduction results in a deeper understanding, in all cases it unifies fields of inquiry. Thus the identification of light with electromagnetic radiation gave us a deeper understanding of light and brought about the unification of two previously separate fields of inquiry. The identification of mental states with brain states of a certain kind is producing not only understanding but also the merger of psychology with neurophysiology, and of psychiatry with neurology. And the identification of historical processes with social changes elucidates history and invites the rapprochement of all the social sciences. Far from impoverishing, such unifications enrich both the reducing and the reduced field of inquiry. However, as we shall see in the next section, integration is no less valuable a device for attaining the unity of knowledge.

Let us start with concept reduction. Table 10.1 exhibits a handful of famous cases of reduction in modern factual science. Every one of them can be understood as a case of concept reduction, which in turn is reducible to a special case of definition. Indeed we stipulate that, if concepts A and B belong to contexts (e.g. theories) $\mathcal{C}_A$ and $\mathcal{C}_B$ respectively, then A is *reducible to* B *in* $\mathcal{C}_B$ if and only if A is definable in terms of B and other concepts of $\mathcal{C}_B$. We call such definition a *reductive definition*, or simply

TABLE 10.1
Some illustrious cases of concept reduction

Reduced Concept	*Reducing Concept*
Heat	Random atomic or molecular motion
Light	Electromagnetic radiation
Chemical reaction	Atomic or molecular combination, dissociation, or substitution
Gene	Segment of DNA molecule
Metabolism	Sequence of biochemical reactions
Goal directed behavior	Negative feedback controlled process
Mental state	State of a plastic neuronal system
History	Process of social change

(concept) *reduction*. Thus the first line of Table 10.1 reads: "The concept of heat (in thermodynamics) is reduced to the concept of random atomic or molecular motion (in statistical mechanics)". Or, to put it in the linguistic mode: "The expressions 'heat' and 'random atomic or molecular motion' have the same denotation in the language of statistical mechanics." Every reductive definition identifies concepts (not things or attributes) that had been treated separately before. This explains its being commonly called a 'bridge formula'.

Our regarding bridge formulas as definitions, hence as identities (Vol. 1, Ch. 10), does not entail that they are arbitrary or conventional. They are definitions *in the new context* (the reducing one). But when first proposed they are *hypotheses* and, as such, they have to win acceptance on theoretical or empirical grounds. (Many an attempt at reduction has failed: we honor only success.) But once the more powerful body of knowledge has passed the tests, the bridge formulas, which were initially hypotheses, are incorporated as intratheoretical definitions. Thus, regardless of the way "light ray" was defined before, one defines it now as the normal to the wave fronts of an electromagnetic radiation field.

The above discussion sheds light on the current controversy over the status of the bridge formulas. Some philosophers admit that they are identities but call them 'contingent identities' (Kripke, 1971); others regard them as identities of things or properties (Causey, 1977); finally others regard them as hypotheses and even laws (Nagel, 1961). The former reason that the reductive definitions, unlike the necessary identities of logic, may not hold in all possible worlds. It must be supposed they have privileged information about such imaginary worlds; in particular they must know that logic holds in all of them, whereas in some of them heat is not random atomic or molecular motion. Having no access to such wondrous information I shall regard reductive definitions as strict identities of concepts. They cannot assert the identity of things or of properties because, if two factual items are the same, then they are one.

As for the view that bridge formulas are laws on a par with laws of motion and others, it may have been suggested by the way they are originally introduced—namely as hypotheses—not by the role they perform in mature theories. Surely it is not a law of nature that the internal pressure of a gas is identical with the average molecular momentum transfer; nor is it a law of nature that combustion is fast oxydation. These are concept identifications: they have no ontic counterpart and have therefore a purely epistemological status. (An omniscient being would dispense with reductive definitions.) To put it into the linguistic mode: a reductive definition effects the translation of an expression of one language into another.

The reduction of propositions follows from the reduction of concepts. Thus "We inherit genes" is reducible to "We inherit segments of DNA molecules". Truth values are preserved under such transformations. Not so meanings, for meaning is contextual (Vol. 2, Ch. 7). Thus the sense of "heat" in classical thermodynamics is the set of all its logical relatives, which is not the same as the set of the logical relatives of "random atomic or molecular motion" in statistical mechanics.

The reducibility of theories (hypothetico-deductive systems) is more complicated. To begin with we distinguish reduction from restriction. Thus the restriction of classical mechanics to point particles is particle mechanics. The latter is a subtheory of the former, obtained by setting the mass density equal to a point-like distribution (a delta), and by discarding the stress tensor and the boundary conditions. (Cf. Bunge, 1967c.) The general definition is this: Let $\mathcal{T}_1$ be a theory with reference class or domain R. Then $\mathcal{T}_2$ is the *restriction* of $\mathcal{T}_1$ to S, where S is a proper subset of R, iff $\mathcal{T}_2$ is a subtheory of $\mathcal{T}_1$ and refers only to the members of S. (For the concept of subtheory see Ch. 9, Section 1.2.) Unlike reduction, restriction does not involve reductive definitions. (Restriction can be characterized in a formal way. Let Q_1 and Q_2 be the sets of basic or undefined predicates of $\mathcal{T}_1$ and $\mathcal{T}_2$ respectively. Then (*a*) Q_2 is a proper subset of Q_1—i.e. some predicates in $\mathcal{T}_1$ have no counterpart in $\mathcal{T}_2$; (*b*) every function in Q_2 is the restriction of the corresponding function in Q_1 to the reference class S of

$\mathcal{T}_2$, or its value in some limit. Thus by setting $c = \infty$ in a relativistic theory, or $h = 0$ in a quantum theory, the corresponding classical theory results.)

Unlike theory restriction, theory reduction involves reductive definitions acting as bridges between the new or reducing theory, on the one hand, and the old or reduced theory on the other. In the most interesting cases, usually ignored in the philosophical literature, reduction involves also additional assumptions not contained in the reducing theory. For example, acoustics is reducible to the theory of elasticity with the help of the reductive definition of "sound" as "wave propagating in an elastic medium". On the other hand the reduction of ray acoustics to general (or wave) acoustics requires the assumption that the wavelengths be small enough so that diffraction is negligible. (The relation between wave and ray optics is parallel.) Likewise the law of ideal gases ($pV = kT$) is deducible from the theory of particle mechanics enriched not only with reductive definitions of "p" and "T" but also with the assumption of molecular chaos. (Incidentally, such reduction has a bonus: it interprets in atomistic terms the phenomenological constant occurring in the phenomenological law.) We shall speak therefore of two kinds of reduction: *strong* or full, and *weak* or partial. Whereas strong theory reduction involves only enriching the reducing theory with reductive definitions, weak theory reduction requires, in addition, subsidiary hypotheses not contained in the reducing theory.

(We summarize the preceding discussion in the following definition. Let $\mathcal{T}_1$ and $\mathcal{T}_2$ be two theories with overlapping reference classes, D a nonempty set of reductive definitions, and A a nonempty set of hypotheses not included in either $\mathcal{T}_1$ or $\mathcal{T}_2$ but couched in the language of $\mathcal{T}_1$. Then (*a*) $\mathcal{T}_2$ is *fully* (or *strongly*) *reducible* to $\mathcal{T}_1$ iff the union of $\mathcal{T}_1$ and D entails $\mathcal{T}_2$; and (*b*) $\mathcal{T}_2$ is *partially* (or *weakly*) reducible to $\mathcal{T}_1$ iff $\mathcal{T}_2$ follows logically from the union of $\mathcal{T}_1$, D, and A. Moreover we shall say that $\mathcal{T}_1$ *helps explain* $\mathcal{T}_2$ iff $\mathcal{T}_1$ is mechanismic and $\mathcal{T}_2$ is phenomenological, and $\mathcal{T}_2$ is fully or partially reducible to $\mathcal{T}_1$.)

Theory restriction is straightforward. (However, it requires more than mathematical competence, namely the knowledge and intuition necessary to choose intelligently what to restrict.) Theory reduction is far more complicated: it requires familiarity with the two theories concerned, as well as an ability to grasp analogies between facts of different appearance. And partial reduction requires, in addition, the imagination that goes into the invention of the suitable additional hypotheses. Think of the ingenuity that went into the construction of modern solid state theory, which is based on quantum mechanics but is neither a subtheory nor a restriction of the latter. More on this below.

The most interesting philosophical feature of reduction is that it explains the emergence of novelty, i.e. the formation of things possessing properties not shared by their components. Scientists are often suspicious of the concept of emergence for regarding it as obscure. However, it can be defined in exact terms, namely as follows. Let s be a system with A-composition, i.e. the components of which (at a given level of analysis) are of type A, where A is a natural kind such as the collection of all atoms, or cells, or societies. If P is a property of s, then P is *emergent* with respect to A if, and only if, no component of type A possesses P; otherwise P is a *resultant* property with respect to A. *Example*: Neural connectivity is not a property of individual neurons but of neuronal systems. And mental activity is not a process in an individual neuron but one in highly complex neuronal systems with variable connectivity. Hence a neurophysiological explanation of perception, memory, or intelligence, will require not only knowledge of individual neurons but also data and hypotheses concerning the interactions among neurons to produce neuronal systems (psychons) endowed with the emergent properties which we call 'mental'. In short, sometimes emergence can be explained by reduction, i.e. in terms of lower level entities and their interactions.

Emergence is not exclusive of the higher levels of organization of matter but occurs on all levels. Consider three things, A, B, and C, of different kinds and capable of assembling into three component systems. In principle six different linear systems are possible, among them ABC and BCA. The structure (order, configuration, pattern of interactions) of the system will certainly be determined by the interactions among its components. But the structure of a system is an emergent property of it, one that neither of its components possesses, and therefore one that could not have been predicted from a knowledge of the components alone.

Consider two specific examples, and in the first place the formation of one of the simplest systems we know of, namely a hydrogen atom. It is composed of a proton and an electron interacting via an electric field. To a first approximation we may disregard its environment, i.e. we may assume that it is placed in a perfect dark vacuum. The application of quantum mechanics to the study of this system explains the emergent as well as the resultant properties of this system. (The explanatory schema is that of a partial reduction: quantum mechanics, enriched with a set of subsidiary assumptions concerning the composition and structure of the system, entails the explanandum. The resulting conceptual system is a bound theoretical model in the sense of Ch. 9, Section 1.2.) The resultant properties of the system are the total mass, charge, angular momentum, and spin. The emergent properties are the discrete energy spectrum (a denumerable infinity of energy states), the probabilities of radiative decay and excitation, and the corresponding probabilities of transition among the various energy levels. None of these properties is possessed by either of the components of the system.

One of the most spectacular successes of the reductionist program is contemporary solid state physics. Consider a length of copper wire. Every one of the atoms that composes it can be accounted for, in principle, by the quantum theory of atoms. However, the body composed of these atoms has bulk properties, such as electrical and thermal conductivity, malleability, and brilliance, which are emergent with respect to the component atoms. These properties are not represented in the quantum theory of atoms, hence they cannot be explained by the latter without further ado. However, solid state physics explains those emergent properties on the basis of quantum mechanics, namely adding to it certain hypotheses concerning the copper crystal lattice, the electrons wandering through it, and the interactions among the copper ions and the electrons moving about in the lattice. Clearly, then, this reduction is of the partial or weak kind. So is the reduction of genetics to molecular biology.

From an ontological viewpoint there are two kinds of reduction: micro and macro. *Microreduction*, or bottom-up analysis, takes the elementary components for granted and attempts to explain the whole in terms of the former and their mutual actions. On the other hand *macroreduction*, or top-down analysis, takes the whole for granted and tries to explain the behavior of its parts in terms of the former. Both accounts have been known from Antiquity: microreduction to the Greek and Indian atomists, macroreduction to Aristotle. Of the two reductionist programs microreductionism has proved by far the more fertile.

The reasons for the microreduction (or atomistic) program are three. One is the ontological hypothesis that all macrosystems are composed of microphysical things such as atoms. Another reason is epistemological, namely that, if one knows the way the components of a system behave, one should be able to infer how the system as a whole behaves. The third reason is methodological: the laws of the microcomponents are supposed to be simpler, hence more readily known, than the laws of the systems they compose. The first reason is legitimate, but we have known this only since the beginning of our century. The second reason holds only in part, for in order to account for the emergent properties of a whole we need to know not only its components but also the mode of their composition (i.e. the system structure) as well as its environment. And the third reason turned out to be false: in fact most known macrolaws are simpler than the microlaws. (Think of the quantum theories *vis-à-vis* their classical counterparts.) The reasons for this greater simplicity of macrolaws are the following: (*a*) a system behaves as a unit in certain respects; (*b*) some of the random fluctuations in the components either cancel out or are unimportant at the macrolevel, and (*c*) the system can be freer than its components, every one of which interacts strongly with other components—for otherwise there would be no system.

The first of the above-mentioned reasons suffices to explain the sensational success of the microreduction program. However, in many cases microreduction is insufficient and must be supplemented with the macroreduction perspective. Thus to explain the peculiar properties of the surface of a liquid we need to know that it is the boundary of a system; to explain the activities of the motor cortex we need to know that it controls the limbs; to explain a machine we need not only a knowledge of its parts but also of its global functions. In all these cases a synthesis of micro and macroreduction is called for.

Theory reduction is one way of explaining emergence. (It is not the only way: in some cases one succeeds in building from scratch a theory explaining the formation of systems of some kind.) However, philosophers concerned with preserving the marvellous variety of reality fear that reduction may eliminate emergence. This fear betrays uncertainty about emergence itself and misunderstanding about explanation. The explanation of a fact does not eliminate it. (What we do eliminate sometimes is nonfacts, i.e. false conjectures about nonexistent facts, as when we explain the fact that two people happened to think of the same object at the same time as a coincidence instead of as evidence for telepathy.) Far from denying the existence of different levels of organization, reduction relates them: i.e. it helps understand the level structure of reality instead of leveling it down or up (Bunge, 1977b).

In any event, the reduction operation has a number of aspects in addition to the logical one, which is almost exclusively that studied in the philosophical literature. See Table 10.2.

How far can or should reduction be pushed? Three different answers to this question have been proposed: antireductionism, radical reductionism,

TABLE 10.2

The various aspects of theory reduction. $\mathcal{T}_1$ = reducing theory, $\mathcal{T}_2$ = reduced theory, D = set of reductive definitions, A = set of additional promises (data or hypotheses)

Aspect	*Full or strong reduction*	*Partial or weak reduction*
Logical	$\mathcal{T}_2$ follows from $\mathcal{T}_1$ plus D.	$\mathcal{T}_2$ follows from $\mathcal{T}_1$ plus D and A.
Semantical	Every predicate in $\mathcal{T}_2$ is in $\mathcal{T}_1$, possibly via D. (So all the peculiar predicates of $\mathcal{T}_2$ are eliminated.)	Some predicates in $\mathcal{T}_2$ are not in $\mathcal{T}_1$ but are built in terms of predicates in $\mathcal{T}_1$, D, and A.
Ontological	The reference class of $\mathcal{T}_2$ is included in that of $\mathcal{T}_1$.	Some referents of $\mathcal{T}_2$ are emergent with respect to those of $\mathcal{T}_1$.
Epistemological	$\mathcal{T}_1$ helps explain the facts accounted for by $\mathcal{T}_2$.	$\mathcal{T}_1$ helps explain the facts accounted for by $\mathcal{T}_2$, which in turn explains emergents w.r.t. $\mathcal{T}_1$.
Methodological	$\mathcal{T}_2$ is a motivation and touchstone for $\mathcal{T}_1$. $\mathcal{T}_1$ may force some changes in $\mathcal{T}_2$.	$\mathcal{T}_1$ is a tool for building $\mathcal{T}_2$. Confirmation of $\mathcal{T}_2$ reinforces $\mathcal{T}_1$, but refutation of $\mathcal{T}_2$ may be blamed on A or D.
Pragmatical	Some computations are simpler in $\mathcal{T}_2$ than in $\mathcal{T}_1$.	Some computations feasible in $\mathcal{T}_2$ are impossible in $\mathcal{T}_1$ (for want of A's).
Historical	Usually $\mathcal{T}_2$ precedes $\mathcal{T}_1$ in time and helps set up the semantic assumptions of $\mathcal{T}_1$.	Usually $\mathcal{T}_1$ precedes $\mathcal{T}_2$ in time and serves as a foil for building $\mathcal{T}_2$.

and moderate reductionism. *Antireductionism* rejects any attempt to understand facts on one level of organization in terms of hypotheses and data concerning some other levels, on the ground that "the whole is more than the sum of its parts", a popular slogan presumably meaning that wholes possess properties not shared by all its components. The ontological counterpart of antireductionism is *holism*, or the view according to which reality is composed of wholes whose properties are independent of those of their components, which are subordinated to the totalities. The antireductionist strategy is to concentrate on wholes in the hope of understanding the nature of their components. (*Examples*: studying whole organisms and whole institutions independently of the study of their components.) Antireductionism, still found in the backwaters of science and the humanities, is refuted every time a chemical process is explained with the help of physical theories, every time a biological function is explained with the help of physics and chemistry, and every time a social process is explained in terms of the needs, beliefs and aspirations of individuals.

Radical reductionism claims that all concepts, hypotheses and theories concerning things on a given level are reducible to those referring to things belonging to some other levels. The ontological companion of radical reductionism is *physicalism* (vulgar materialism), according to which every thing is a physical entity, and the differences we draw between the physical, the chemical, the biological, and the social, are merely differences in complexity. The radical reductionist strategy is to concentrate on the components of a system in the hope of understanding the latter. (*Examples*: studying cells before whole organisms, and individuals before social systems.) Although in the vast majority of cases reduction is partial rather than full, radical reductionism is extremely popular among natural scientists.

Radical reduction has failed even in physics. In fact physicists currently admit that there are at least three fundamental distinct theories, none of which is reducible to the other: quantum mechanics, quantum electrodynamics, and the relativistic theory of gravitation. (The status of chromodynamics is still uncertain.) These theories are mutually irreducible because they refer to things of radically different kinds: particles, electromagnetic fields, and gravitational fields. (They can of course be mutually related but not reduced to one another.) Classical mechanics has not yet been successfully reduced to quantum mechanics: it is not known which are the subsidiary hypotheses to be adjoined to quantum mechanics in order to obtain the full theory of the motion of extended bodies, i.e.

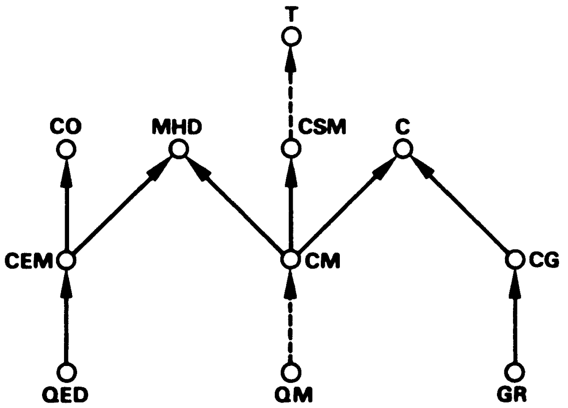


Fig. 10.2. Three basic physical theories of the day: quantum electrodynamics (QED), quantum mechanics (QM), and general relativity or the theory of gravitation (GR). Arrows designate reduction, still incomplete in the case of dashed lines. Classical optics (CO) reduces to classical electromagnetism (CEM), which in turn reduces to quantum electrodynamics (QED). As for QM, one hopes it will be shown to entail classical mechanics (CM), which in turn yields classical statistical mechanics (CSM), which, suitably enriched, should eventually yield thermodynamics (T). Relativistic theory of gravitation (GR) entails the classical theory of gravitation (CG). The union of CEM and CM entails magnetohydrodynamics (MHD), and the union of CM and CG entails celestial mechanics (C). All these reductions are partial. From Bunge (1977b).

continuum mechanics. On the other hand, for strong electromagnetic fields quantum electrodynamics goes over into classical electrodynamics. Likewise, for weak gravitational fields the relativistic theory of gravitation reduces to the classical theory of gravitation. See Figure 10.2. Most of the other reductions cited in the scientific and philosophical literature are instances of wishful thinking. (See Bunge (1973a) for details on reduction relations in physics.) The positivist dream of building a unified science by successive microreductions (Neurath, 1938; Oppenheim and Putnam, 1958; Causey, 1977) has turned to be just that: a dream not even fulfilled in physics. But at least it has been an inspiring dream unlike the antireductionist nightmare. Besides, the unity of science and, indeed, of all genuine knowledge, is fact not fancy: see Ch. 14, Section 3.

Finally, *moderate reductionism* is the strategy of reducing whatever can be reduced (fully or partially) without ignoring variety and emergence—nay, aiming at accounting for them. The ontological counterpart of moderate reductionism is *emergentism* (see Bunge, 1981a), according to which wholes have properties not shared by their components but which, far from being self-existent, result from the latter. The moderate reductionist strategy is to study the whole as well as its parts, in the hope

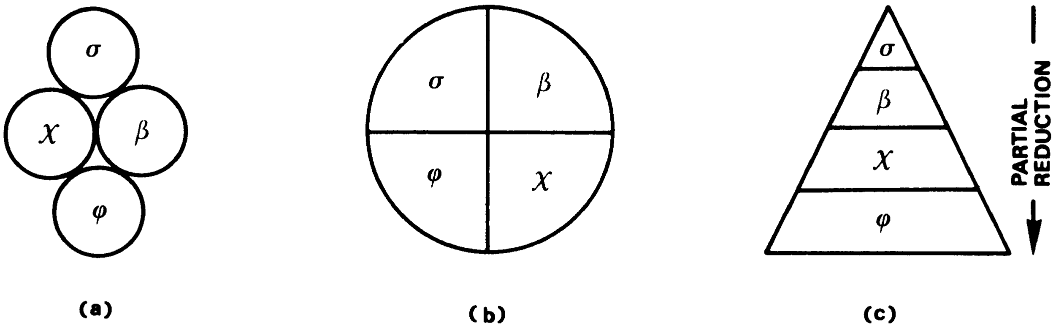


Fig. 10.3. The world and its knowledge. φ = Physics, χ = Chemistry, β = Biology, σ = Social Science. (*a*) Antireductionism: unrelated compartments. (*b*) Radical reductionism: different sectors of a single whole. (*c*) Moderate reductionism: each level higher than the physical one is rooted to the previous ones.

that each kind of research will supplement and spur the other. (*Examples*: studying both individual neurons and neuronal systems, persons and social systems.) Properly analyzed, the most important cases of reduction have been instances of partial reduction. Such successes have refuted both antireductionism and radical reductionism.

Contrary to antireductionism, the various fields of research are becoming more and more integrated, not only by the use of a common methodology underlying the diversity in techniques, but also by relations of (partial) reduction as well as by mergers. And, contrary to radical reductionism, an increasing number of research fields are revealing that the world, though one, has a marvellous variety (accounted for by an increasing number of theories) and a level structure (somehow mirrored in the hierarchy of the sciences). See Figure 10.3. In sum, moderate reductionism is a progressive and realistic strategy, whereas antireductionism is regressive, and radical reductionism utopian.

3.2. *Integration*

Things cannot always be explained by microreduction: quite often they can be explained only by placing them in a wider context. See Figure 10.4. In turn the consideration of such a wider context requires bringing together or consolidating a number of scraps of information and conjecture dispersed among several fields of inquiry. Occasionally such an operation of integration or synthesis brings about the merger of theories and even entire disciplines.

For example, one does not explain the meaning of a word by analyzing it

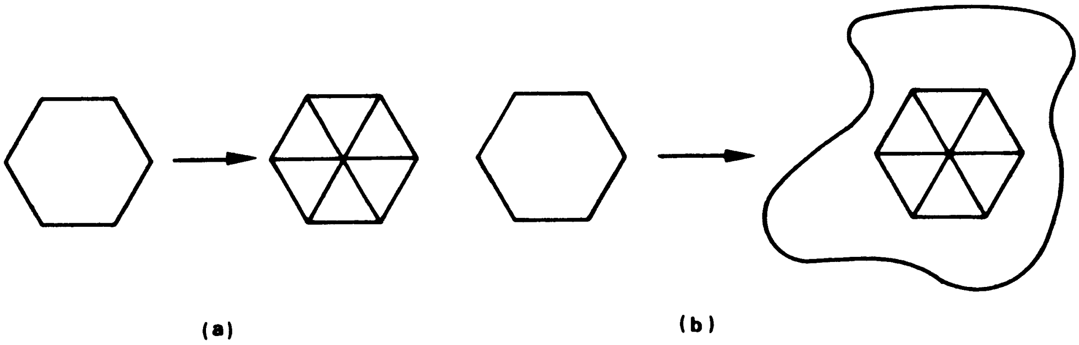


Fig. 10.4. (*a*) Analysis through reduction. (*b*) Analysis through integration.

into its component letters, or even phonemes, but by relating it to other words. The life history of an individual organism is explained not only in terms of genetics and physiology but also in ecological terms. In particular, the behavior of a psychologist could not be explained by only looking into his brain or by watching some indicators of his brain activity: one must also study the systems of which he is a component, such as his family, his laboratory, his scientific community, and even his entire culture. A last example: classical microeconomics, a typically reductionistic enterprise, has failed to explain the behavior of households and enterprises for disregarding their wider social context. One cannot account for the actions of a thing apart from the state of the system it is embedded in. Thus one could not possibly understand why the manager of a firm decides to manufacture product X unless there is a market for X (or one can be created by advertising). And this depends not only on individual needs and wants but also on macroeconomic circumstances and, indeed, on a number of political and cultural variables as well. In sum, what we need is the integration of microeconomics and macroeconomics, not the reduction of one to the other. (More on this in Vol. 6.)

The integration of approaches, data, hypotheses, theories, and even entire fields of research is needed not only to account for things that interact strongly with their environment. Epistemic integration is needed everywhere because there are no perfectly isolated things, because every property is related to other properties, and because every thing is a system or a component of some system (Vol. 4). Thus, just as the variety of reality requires a multitude of disciplines, so the integration of the latter is necessitated by the unity of reality. See Figure 10.3.

Because the unification of knowledge via integration has been sadly neglected by philosophers, particularly those of reductionist leanings, it will be convenient to draw a somewhat long list of successful integrations in science. (*a*) Until recently physicists tried to build one theory per particle species. The large number of particle species, as well as the many relations among them, has necessitated an integrative approach based on the conjecture that each member of a particle species is a state of an underlying particle, much as the electron and the positron are two states of a single particle. (*b*) The relativistic theory of gravitation is not self-contained: in order to solve its equations one must borrow the so-called constitutive relations (determining the so-called matter tensor) from other fields of inquiry. (*c*) The solid state physicist does not start from scratch but makes use of a number of theories, from quantum mechanics to statistical mechanics. (*d*) The theoretical chemist can rarely afford to make *ab initio* calculations: usually he uses both classical chemistry (in particular chemical kinetics) and quantum mechanics, and sometimes even classical particle mechanics. (*e*) Until recently the formation of mountains, earthquakes and vulcanism were treated as separate processes and even by different disciplines. Ever since the plate tectonics revolution, geology treats them as different aspects of the same process. (*f*) Contemporary theoretical meteorology is the outcome of the synthesis of two formerly disjoint theories, namely fluid dynamics and thermodynamics. (The former, which dealt with ideal viscosity-free fluids, had to be rendered more realistic before it could be applied to the atmosphere and the ocean.) (*g*) The stellar atmospheres (in particular the solar corona) could not be studied properly until fluid dynamics and magnetism were fused into magnetohydrodynamics. (*h*) Physiological psychology (or neuropsychology) is a merger of neurophysiology and psychology. Its aim "is not to replace one science by another, but to close the gap between the two sciences—to proclaim their unity, in principle" (Bindra, 1976, p. 19). This unified approach should overcome the one-sidedness of the earlier, reductionist schools, which had attempted to explain behavior, affect, perception, and cognition, in terms of either external stimulation (behaviorism), affect (psychoanalysis), cognition (cognitivism), or some other alleged first mover.

We are now ready to analyze the merger or amalgamation of two or more theories (hypothetico-deductive systems). During the discussion it will be convenient to bear in mind such paragons of theory merger as analytic geometry (the synthesis of synthetic geometry and algebra), celestial mechanics (the union of mechanics and the theory of gravitation), electromagnetic theory (the merger of the theories of electricity and magnetism), thermomechanics (the synthesis of continuum mechanics and thermodynamics), the synthetic theory of evolution (the merger of Darwin's classical theory of evolution and genetics), and evolutionary ecology (the currently attempted merger of population genetics and population ecology).

The simplest case of theory merger is the union of two or more theories (or rather their sets of formulas). Obviously, not every such union is "meaningful". The precursor theories must share referents and therefore also some specific concepts (variables, functions). This is the case with mechanics and the theory of gravitation, the precursors of celestial mechanics. Here the force that occurs in Newton's laws of motion is calculated in the theory of gravitation. (The latter theory did not exist in Newton's time: it was a 19th century creation.) Likewise, the synthetic theory of evolution is the union of Darwin's theory with genetics. However, there are not many more examples of theory union.

In the vast majority of cases the precursor or founding theories have to be supplemented with formulas connecting concepts of the two theories and thus acting as a glue between them. For example, an algebra with a $+$ operation and another with a $\times$ operation can be amalgamated only if both operations are defined on the same set (i.e. if the referents are the same) and if glue formulas such as "$(x + y) \times z = x \times z + y \times z$" are added. To take an even more telling example, and the pioneer in theory merger: Descartes built analytic geometry by assigning every point in Euclidean n-space a unique n-tuple of real numbers and conversely (i.e. by postulating a bijection between E^n and $\mathbb{R}^n$), thus mapping every geometric figure onto a set of n-tuples of real numbers. The theories of electricity and magnetism could not have been synthesized into the electromagnetic theory without further ado: Faraday's law of induction and Maxwell's conjecture of displacement currents had to be added.

In general we shall say that $\mathcal{T}$ is a *merger* of $\mathcal{T}_1$ and $\mathcal{T}_2$ if, and only if,

(i) $\mathcal{T}_1$ and $\mathcal{T}_2$ share some referents and some concepts (e.g. functions);

(ii) there is a (possibly empty) set G of (glue) formulas relating some concepts of $\mathcal{T}_1$ to concepts of $\mathcal{T}_2$, and

(iii) the (glue) formulas in G are sufficiently confirmed.

Condition (i) excludes theories that have nothing to do with one another. (*Caution*: the sharing of spatial or time coordinates does not count, for they are universal features.) Condition (ii) makes room for theory union as the particular case when G is empty. And condition (iii) is added because, in principle, there are infinitely many possible glue formulas. Only those matching the available data will hold the original theories effectively together. By the way, note the methodological difference between glue formulas and the reductive (or bridge) identities of the previous sections: whereas the latter are definitions, the former are postulates.

To sum up, integration is at least as frequent as reduction and just as important a factor of the unity of knowledge. See Figure 10.4. Integration and reduction accompany specialization and make up for its narrowness and centrifugal tendency. Integration is particularly conspicuous in the study of macrosystems, natural or social, because of their multiple aspects and the several levels of organization they cross. Despite their enormous variety, science and technology are one in multiple ways: (*a*) all the chapters of science and technology share a general outlook or philosophical background; (*b*) all the sciences and technologies share logic and mathematics; (*c*) all the sciences and technologies share the scientific method; (*d*) every science and every technology deals with one part or aspect of reality (the only world there is); (*e*) every science and every technology is closely related to at least one other science or technology, so that the totality of sciences and technologies forms a system; (*f*) all of the sciences share the goal of getting to know laws, systematizing them into theories, and using them for cognitive or practical purposes; (*g*) all of the technologies share the goal of using knowledge in the design of artifacts. Therefore there is no need to try and reduce all sciences to physics (as physicalists have demanded) or to psychology (as suggested by phenomenalists), or to merge all theories into a single all-encompassing theory. Science and technology are already one in spite of the inevitable failure of such quixotic attempts. What has to be done is to tighten the system, by reducing the reducible and integrating the integrable.

4. FORECASTING

4.1 *From Expectancy to Forecast*

All higher vertebrates seem capable of expectancy, i.e. of adopting the attitude of anticipating impending events. The animal that expects something to happen may be curious, elated, or frightened, but not indifferent to the future: he has a definite "set" (*Einstellung*). He may expect something to happen or he may have a definite expectation on the strength of a learned association. If he has learned to expect event *B* to follow event *A*, either regularly or with some frequency, a characteristic wave (the *E* wave) will appear in his forebrain (Grey Walter *et al.*, 1964).

The predictive abilities of some animals are quite remarkable. If I throw a stick for my retriever and continue to walk, the animal will meet me further on along the pathway without retracing his steps: somehow he has "computed" my path and adjusted his own to it. When we perceive a regular wave motion we detect its periodicity and anticipate the motion. Similarly we can often complete sentences that other people start and fragments of melodies we have caught. We do all this without using any explicit predicting techniques. Expectancy, and a modicum of forecasting ability, seem to be in the survival kit of all higher vertebrates. *Praemonitus, praemunitus* (Forewarned, forearmed).

Presumably, the more strongly an event is correlated with a previous event, the easier it is to learn that the latter is a predictor of the former. However, such objective correlation is not sufficient for learning about it. Contrary to what one might expect intuitively, experiment shows that animals learn to associate an event *A* to a latter and correlated event *B* only if they are *unable* to predict *B* from *A*: the more surprising an event the better we learn to associate it with its predictor (Rescorla and Wagner, 1972). The reason could be that the more surprising events are more intensively "processed" in the brain.

Although the neurophysiology and psychology of expectancy and forecast have made a good start, we still have to conjecture the way we learn predictive relations. An attractive conjecture is the following (Bindra, 1976). Suppose an animal learns that event *A* is often followed by event *B*, and so learns to predict *B* from *A*. What happens in its brain is that the events as well as their relation are represented by activities of neuronal systems which Bindra calls 'contingency organizations'. If the contingency organization representing event *A* is activated in the animal that has learned the *A*–*B* connection, then it stimulates or primes the perceptual representation of *B* even before *B* happens. That is, the psychon of the *A*–*B* connection causes an advance excitation of the psychon of the predicted event *B*, or at least predisposes the latter to activation (by lowering its excitation threshold). The actual perception of *B* is therefore facilitated by the predictor event. In other words, the animal that knows that *A* is followed by *B* expects *B* upon perceiving *A* (or imagines *B* upon imagining *A*) and is thus better prepared to perceive *B* and act in consequence. See Figure 10.5.

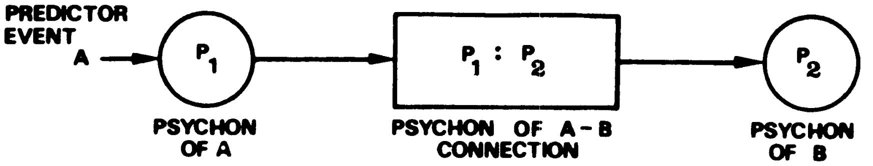


Fig. 10.5. The neurophysiology of prediction. Predictor even A activates psychon p_1, which acts on contingency organization $p_1 : p_2$, which in turn controls the excitation level of p_2, the psychon of the predicted event B. From Bindra (1976, p. 132).

The fact that we have definite neural mechanisms for prediction does not mean that all our forecasts are correct. Many things may go wrong with our predictive mechanisms. We may have learned the incorrect pairings, e.g. to expect a catastrophe every time our path crosses a black cat. Or we may have learned the correct pairings but the predictive mechanism was inhibited, or was interfered with by some other psychon. Or, and this is the most common case, we may have failed to learn any predictive pairings at all—and yet we may be foolish or hard pressed enough to make predictions. This explains the well-known fact that most intuitive forecasts, and even a high fraction of expert forecasts (in medicine, business, and politics) are erroneous.

We hardly needed psychological research to tell us this. What psychologists can do is to try and find out what prediction strategies people use and how they perform. Kahneman and Tversky (1973) studied the problem and found that their subjects, far from using all the evidence at their disposal, selected evidence of one kind only and, moreover, ignored evidence that, if attended to, would have elicited a different judgment. (Specifically, the subjects were asked to predict the area of specialization that a certain individual, whose personality sketch had been drawn on the basis of projective techniques, and which was known to them, was likely to choose. The subjects relied only on this personality sketch, even though knowing that projective techniques are unreliable; and they predicted, e.g. that a given subject would choose computer science, even though knowing that the fraction of the student population that chooses computer science is small.)

From a logical point of view forecasts are special cases of description. They describe events or processes that are not yet or may never be, and therefore are conjectural rather than known with certainty. Indeed a description may refer to past, present, or future events. A description of future events is called a *prediction*, one of past events a *postdiction*. Predictions, postdictions and descriptions of current events are all nonexplanatory accounts. True, some predictions and postdictions are more chancy than some descriptions, but this is not always the case. Thus, a prediction of planetary movements is more accurate than either a postdiction or a prediction of human behavior. In sum, prediction and description do not differ in logic and they need not differ in truth value. The difference between them is methodological: the ultimate check of factual hypotheses and theories consists in checking the predictions or postdictions made with their help. But checking belongs in Ch. 11.

Forecasts can be classed into intuitive and rational. An *intuitive forecast* uses only tacit premises or perhaps none; a *rational forecast* employs only explicit premises: it is the logical consequence of a deductive argument from hypotheses (e.g. law statements) and data. Intuitive forecasts are more or less imaginative inventions, and often just wishful or dreadful thinking. Examples of this kind of forecast are the long term weather forecasts that the *Farmer's Almanac* has been publishing every year since 1792 with the help of a "secret formula". Though widely believed, such forecasts have been shown to be not much better than those obtained by flipping a coin (Walsh and Allen, 1981). Because the premises of an intuitive forecast are hidden or nonexistent, we cannot learn anything when watching its performance. On the other hand the success or failure of a rational forecast is likely to teach us something about the relevance, accuracy, and sufficiency of the premises employed in formulating it.

Intuitive forecasts on a grand scale are called *prophecies*. In the past prophesying was an occupation of shamans, priests, astrologers, soothsayers, and politicians. It has now become an industry that may be called *prophetic futurology* (or *pop forecasting*), to distinguish it from serious futurology, which is concerned with planning on a grand scale, i.e. with redesigning society. Unlike the prophets of the Old Testament, who were interested mainly in domestic events and knew no electronic gadgets, the high priests of modern prophecy deal with world events and extrapolate from modern technology. For instance, they prophesy that before long everyone will be assured of a decent income and will enjoy programmed dreams and individual flying platforms (Herman Kahn). They tell us matter-of-factly that the 21st century will be run by computers that (who?) will dispense justice and keep us out of mischief, will solve all intellectual and practical problems, will keep track of all we do and fill in for us when we fail to do our duty, on top of which they will design and construct even more intelligent and efficient machines. Sadly, none of the modern prophets were able to forecast any of the major events of the recent past, such as the oil crisis, stagflation, the Green Revolution, the Islamic revival, the defeat of the U.S.A. in Vietnam, the Sino–American alliance, the ousting of Somoza, the return of Juan Perón and Indira Gandhi to power, the Falklands war, or the decline in the pace of technological innovation.

Modern prophecies, just like those of Ezequiel, Daniel, and Isaiah, are of the form *B will happen at time C*—e.g. money will have disappeared by the year 2000. That is, they are unconditional statements, unlike the conditionals of scientific prediction, which are of the form "If *A* happens (or continues to be the case), then *B* will (or may) be the case at time *C*". But, unlike the prophets of old, the modern prophets make wild extrapolations from current technological trends and they disregard the social constraints: they take it for granted that whatever the technologist may come up with will instantly be adopted across the frontiers. Moreover, they assume wrongly that they have sufficient knowledge of the present, which is never the case if only because they are usually specialists with a one-sided vision of the world.

Some of the pitfalls of intuitive forecasting can be avoided by pooling and sifting the opinions and speculations of a number of experts in a given area and determining the points of consensus. This is what the Delphi "method" boils down to—although it was initially advertised as "a tested technique of long-range forecasting" (Helmer, 1966). It is now regarded as merely a procedure for gathering "soft" (i.e. subjective and unchecked) data in the social technologies (Linstone and Turoff, 1975). It lacks theoretical justification and it is yet to be seen whether it is successful.

Rational forecasts, like descriptions, can be extratheoretical or theoretical. Most of the rational forecasts made in medicine, education, economics, and politics, are extratheoretical: they employ predictors, not theories. For example, the general practitioner prognosticates on the basis of vital indicators, the teacher and the businessman on the strength of past performance, and the economist and politologist extrapolate present trends. Any sign or indicator used to make forecasts is called a *predictor*. For example, we employ the results of entrance examinations as predictors of scholastic performance (even though we ought to know that they are poor predictors), on the assumption that the student will continue to behave as in the past. This is the way high school students are admitted to college, graduates hired as professors, employees promoted to management, and managers to directors. The hypothesis that anyone who is good at any given level must be good at the next higher level works much better than the superficial impressions one may get from an interview, and better than recommendations. However, it is not quite correct, because the abilities required in each case are different. Oversight of this well known fact leads to adopting the famous Peter Principle: "Everyone gets promoted to his own level of incompetence".

(One way of elucidating the notion of a predictor is as follows. Suppose we know two kinds of events, E_1 and E_2, and know also that they are related. This relation can be either constant (e.g. causal) or probabilistic. I.e., either some event of kind E_2 happens every time an event of type E_1 occurs, or there is a definite probability of an E_2 happening every time an E_1 occurs. In the first case any functional relation between some property of E_1's and some property of E_2's can be used as a predictor of E_2's. In case the relation is stochastic we may take the difference between the conditional probability $Pr(e_2|e_1)$, for e_1 in E_1 and e_2 in E_2, and the absolute probability $Pr(e_1)$, to measure the value of E_1 as a predictor of E_2. More precisely, we may define the *predictive value* of E_1 for E_2 as

$$v(e_1, e_2) = Pr(e_2|e_1) - Pr(e_1) \quad \text{for all } e_1 \in E_1 \text{ and } e_2 \in E_2.$$

This difference is positive only in case e_1 is favourable to the occurrence of e_2; it vanishes if e_1 and e_2 are statistically independent. If the predictive value of E_1 for E_2 is positive, E_1 can be said to be a *predictor* of E_2. From the above we see that a predictor may be a cause, an indicator of a cause, or any event functionally associated with the event of interest.)

All governmental agencies and many private enterprises extrapolate current trends in population, energy consumption, industrial production, use of various items, and so forth. Such extrapolations assume that nothing is done to curb the given tendencies. Such forecasts, which often involve massive use of data and computers, are sometimes called *scenarios*. Scenarios have little cognitive value because they result from mere extrapolation of current trends. But, by the same token, they may have immense practical value in sounding timely alarms and suggesting plans. For example, one of them tells us that, if the world population were to continue growing at its present rate, there would be no standing room left by the end of the next century; another that, if car sales continue at the present rate, every human, including infants, would own a car by the year 2050; still another, that the consumption of hard drugs would exceed that of food.

Because they are absurd, such scenarios suggest that something is wrong with some of our present systems: that they will break down unless we interfere with them. "By so doing one can plan interventions *now* rather than wait, as is usually the case, until the system is in a state of crisis" (Ackoff, 1978, p. 128). For example, all experts in nuclear warfare are agreed on the following: (*a*) in case of a (or rather *the* first and last) nuclear war, all the major towns in the northern hemisphere would be destroyed; (*b*) there is no effective shelter against nuclear blast or even against radioactive fallout; (*c*) as long as there are stockpiles of nuclear warheads it is possible that a nuclear holocaust be started by an irreflexive head of state or even by mistake. Such forecasts should suffice to persuade any rational being that total nuclear disarmament is necessary (though not sufficient) to preserve the human species, at least for a while.

The rate of success of extrapolations of current trends is far superior to that of intuitive forecasts, such as those based on clinical judgement, unstructured interviews, or projective tests. In particular, statistical forecasts, made on the basis of regression equations, work reasonably well at least in the short run, not only in the social sciences, especially in economics, but also in medicine and clinical psychology (Sawyer, 1966; Dawes, 1980). Even the simplest statistical model (linear regression or correlation) does better than the intuitive forecast. Still, statistical predictions (i.e. extrapolations of present tendencies) are far poorer than theoretical predictions, i.e. forecasts with the help of theories. The reason is that, unlike every other kind of forecast, theoretical prediction involves laws (and also rules in the case of social science), not just trends that can change overnight.

Thus a knowledge of statics, and the information that a particular lever is in equilibrium, allow us to predict that, if the weight or the length of one of its arms changes, the lever will become unbalanced. This forecast is qualitative both with regard to the imbalance condition and the time. An instance of a quantitative forecast is this. If the law of growth (or decay) of a certain magnitude (e.g. the number of individuals in a population) is exponential, and one knows or assumes both the initial value of the magnitude and its growth (or decay) rate, then one can forecast with precision the value of that magnitude at any time. One can also make several other predictions, e.g. the precise time at which the given magnitude will have doubled (or halved). To be sure, such a precise forecast may turn out to be false, but this failure will be instructive: it may indicate that some limiting factor (e.g. scarcity of resources, crowding, or competition) is at play. No such lesson can be learned from an intuitive forecast.

The truth of a forecast must be distinguished from its precision, for—as every fortune teller knows—the less definite a forecast the more chances it has of being true. Consider the following forecasts:

(i) Anything may happen.
(ii) Something will turn up to solve the energy crisis.
(iii) Some day man will learn to control society.
(iv) There won't be any fossil fuel left by the year 2100.
(v) The Halley comet will be visible from Earth on April 11, 1986.
(vi) The probability that an atom of species A in state B will emit a photon within time t, is p.

These forecasts are of different degrees of definiteness. The first is the least definite of all; fortunately we may disregard it, for it conflicts with the principle of lawfulness. The second is mere wishful thinking, since very little, aside from rather slow progress on fusion and satellite solar power, is being done to avert the impending catastrophe. The third too is rather indefinite but it is somewhat better grounded than the preceding, because social science is making progress, and planning is gaining world-wide acceptance. Forecast (iv) results from extrapolating current fuel consumption and population trends, so it is both definite and well grounded—though unreliable because those trends may not continue. The last two forecasts are both definite and well grounded. Unlike the preceding ones, they are theoretical predictions. Let us take a closer look at them.

4.2. *Scientific Prediction*

There are only actuals—by definition. But of course possibility is of the very essence of actuality: nothing can be, or cease to be, unless it is really possible to begin with. (For our definition of real possibility in terms of laws see Vol. 3, Ch. 4.) All actuals change and induce change in others. However, as long as such changes do not occur, they are nothing but possibilities. Nothing prevents us from, and everything advises us to, imagining possibles—possible things, states of things, and changes in such states, i.e. events. This is what forecast is all about: imagining possible individuals. Such imagination may be fanciful, as in fiction and prophecy, which are restricted at most by logic; or it may be controlled, as in scientific forecast.

Controlled forecast, as practised in science and technology, is of either of two kinds: *empirical* (though not intuitive) and *theoretical*. The former uses statistical correlations, regression lines, and other statistical tools; on the other hand theoretical forecast involves the knowledge of laws. Evidently, theoretical forecast is in principle superior to empirical forecast, because laws are deeper and firmer than correlations. However, theoretical forecast may score lower than empirical forecast if it relies on incomplete or inaccurate data, as is the case with weather forecasting. Let us catch a glimpse of the latter.

The traditional method of weather forecasting uses empirical generalizations, while the theoretical method employs a sophisticated mathematical model of the atmosphere that combines fluid dynamics with thermodynamics. Use of the former method calls for considerable flair that can be educated. On the other hand, once the theoretical model has been chosen and the data are in, numerical forecasting can be entrusted to a high speed computer (as is done at the U.S. National Meteorological Center and a few other institutions). Unfortunately the meteorological data are not only inaccurate but also incomplete. For one thing they do not include values of the entropy, the energy fluxes, and other functions; for another, they fail to cover vast regions of our planet, such as the oceans and the arctic and equatorial regions. So, even assuming that the theoretical model is correct—a very doubtful assumption, if only because it is a non-probabilistic model—it contains variables that go unmeasured, and others that are measured only in some places and with large errors. As a result, the empirical weather forecasts are still, after four decades of work on theoretical forecasts, more accurate than the theoretical ones.

This *impasse* has led some meteorologists to suggest giving up computer simulations of the atmosphere (Sanders, 1979) or even abandoning the project of theoretical forecast altogether (Ramage, 1976). Yet presumably something can be done about this crisis—e.g. to try more sophisticated models of the atmosphere, and to multiply the number of (automated) meteorological stations. (Incidentally, weathermen are in the habit of speaking of probabilities, e.g. of precipitation. This use of the term 'probability' is incorrect, because neither the empirical regularities nor the theoretical models involved in weather forecasting are probabilistic. When the weatherman learns that in the past p percent of the time the atmosphere looked in such and such a way the next day it rained, he forecasts: "The probability of rain tomorrow is p". Here p is a percentage not a probability. There are no probabilities unless there is randomness: see Bunge, 1981c.)

Empirical (statistical) forecasts are based on unexplained correlations, which are precise measures of the strength of the association between two variables, each of which represents a property of some concrete thing. Thus if we have found that X correlates highly with Y, then every time we find X

we may expect to find Y as well—provided the correlation holds out instead of being spurious as in the famous cases of the correlations between grain crops and sunspots, or between divorces and foxes killed in Finland. For example, for every human population there is a definite correlation between height and weight. This correlation can be used to predict the one from a knowledge of the other in the given population. (But of course it does not apply to a different population or to individuals, who may become leaner or fatter without changing in height.) Once we have obtained a correlation coefficient we can draw a line that allows us to extrapolate into the future—assuming that the correlation will hold out. In particular, we can build autoregressive models centered on autoregressive equations, or finite difference equations for a single variable, e.g. $X_t = a_1 X_{t-1} + a_2 X_{t-2}$, which allows us to forecast the value of X at time t in terms of its values at the earlier times t-1 and t-2. All such models are useful forecasting tools and they abound in the life and social sciences. But it is well understood that they yield soft forecasts far inferior to the ones based on theories.

As we saw in Section 4.1, a theoretical forecast is the conclusion of an argument from two groups of premises: theoretical formulas (in particular law statements) and data. We saw also that the occurrence of the time variable in the premises is sufficient but not necessary to make forecasts. For instance Gibbs' phase "rule" (actually a law statement) states that, for any system in equilibrium, with C independent components, P phases, and F degrees of freedom, $C + 2 = P + F$. Two general predictions follow: (*a*) If either of the three properties of a system in equilibrium changes, the system goes into a nonequilibrium state, and (*b*) if a system is in a nonequilibrium state, it will go over to an equilibrium state if either (or all) of the three properties change in such a way that the phase "rule" is satisfied. These predictions are made on the strength of a law statement that does not contain the time concept, and they require counting but not measuring. An even simpler law statement of the same type is the ecological law that, if there is a relation of mutualism between the members of two species, then if either of them increases in number, the other will increase as well.

Most scientific predictions are forecasts of states, events, or processes of known types. Paragons of accuracy are of course the predictions made with the help of celestial mechanics, quantum mechanics, and quantum electrodynamics. In most cases they forecast not only what will happen (or what cannot possibly happen) but also when the event will happen, or at least the probability that it will happen within the next time unit. (If such probability p is calculated then it can be expected that, on the average, the event concerned will happen pN times out of N occasions.) The standard inference schema of any such forecasts is: *Model of the thing(s) concerned plus data entails prediction*. Since in the cases we have mentioned a general theory is involved, the model is obtained by enriching the former with specific assumptions concerning the composition, structure, and environment of the system of interest. (For the notion of a bound model recall Ch. 9, Section 1.2.)

Some of the most exciting scientific predictions are not of states or events of known kinds but of the very existence of previously unsuspected things. Here is a random sample of the increasing collection of forecasts of this type: Darwin's postdiction of the existence of our hominid ancestors, Maxwell's prediction of electromagnetic waves, Mendeleev's prediction of the existence of the elements scandium, gallium and germanium; Oppenheimer's prediction of neutron stars (pulsars), and the prediction of new, as yet not synthesized molecules, made almost every day by organic chemists. All these predictions and many more were made with the help of scientific systematizations and they were borne out by observation or experiment—nay, they guided the latter. Predictions of the same type, but not yet verified, are those of gravitational waves and of superheavy elements (with nuclei around $Z = 114$).

The most satisfactory theories are those which are at the same time predictive and explanatory. Modern planetary astronomy is of this kind; on the other hand the theory of evolution is far too general to be able to make any quantitative predictions. Ancient planetary astronomy, in particular Ptolemy's, was predictive but not explanatory: it was but a device for computing appearances from appearances. Therefore the ancient skeptics were right in doubting that the sun would "rise" the next day. We can now be certain that it will "rise" because we know the mechanism of the alternation of days and nights, namely the spinning of the Earth. (We may doubt on the other hand whether there will be any life left on earth if the superpowers continue to prepare for the final solution to all our problems.)

How is it possible to predict the existence of unheard-of properties, events, and even things? If every scientific statement were nothing but a datum or an empirical generalization, such predictions would be miraculous. First, because before the actual finding of the predicted object there was nothing to generalize from; second because, on an inductivist philosophy, there is no reason to suppose that properties and things could be necessarily (lawfully) related to one another. On the other hand, on the view that theories are hypothetico-deductive systems, serendipity loses its miraculous quality while keeping its magnificence: if certain assumptions are made, a host of conclusions can be drawn, some of which must be unexpected when only the assumptions are examined. Something similar happened with the predictions made on the basis of classificatory schemata such as the periodic table of the elements and the systematics of elementary particles. In both cases certain pigeonholes appeared empty which arose the curiosity of experimentalists to look for things that could possibly fit into them. These discoveries, then, were not lucky accidents. They only illustrate the power of conceptual systematization (classification and theory) to guide experience and thus enrich it. See Appendix 4 for the notion of predictive power.

According to many thinkers, from Ptolemy and the judges of Galileo to Comte and Lakatos, prediction is the hallmark of science: everything else is unimportant. On the other hand Popper believes explanation to be more valuable than prediction, which he regards as unimportant except for technological purposes. As usual, the truth lies in between: prediction is a mark of science though not its only peculiarity nor characteristic of every bit of it, for none of the extremely general theories can make any prediction unless enriched by subsidiary hypotheses and data. Prediction is valuable in itself because it widens the collection of knowable facts. And it is *a* mark of science because, unless there is some prediction, nothing can be tested, and without testing there is no science. (Pseudoscience, in particular psychoanalysis and parapsychology, makes no precise predictions.)

Can we predict everything? Here again there are three possible answers. Fortune tellers, economic advisors, stock brokers, and some politicians claim that they can predict everything: after all they make a living off prophesying. On the other hand many social scientists and philosophers believe that nothing important can be forecast. As usual, neither the optimists nor the pessimists are right. Predictability is not a property of events but of our cognitive relation with them: we can predict provided we know enough. Moreover, since prediction is performed within a body of knowledge, and every such body is incomplete and inaccurate, not everything can be predicted at a given moment, and if predictable it is predicted with some error. And, granting that we shall never attain perfect knowledge, there will always be unpredictables. But, assuming that future generations will continue to care for scientific research, we can also forecast that knowledge will continue to grow, and thus the number of predictables increase.

Neither the rationalist nor the realist need lose any sleep over unpredictables. What should disturb them is the threat of diminishing research. Still, could we discover any unpredictables, i.e. facts either unknowable or lawless? If unknowable, we need not worry: out of sight, out of mind. If lawless, they would be beyond the reach of science, for lawfulness is a (philosophical) presupposition of scientific research. (This principle is irrefutable: if anyone were to point to a certain fact as lawless, we could ask for some time to investigate it in the hope of discovering its laws.) Two candidates come to mind: time series and the outcomes of research and invention. Let us catch a glimpse of each.

A *time series* is a sequence of data of some kind (e.g. total population, or precipitation) taken at regularly spaced intervals, e.g. years. Whereas some time series are random, others exhibit a pattern. The former are unpredictable unless the underlying mechanism is discovered; the others are predictable within bounds. (The degree of predictability of a time series is measured by the autocorrelation function, whose values are computed from the correlation between observation results separated by two, three, etc., intervals. In a predictable time series the correlation coefficient will, typically, fluctuate at the beginning, settling down to a constant value after a while.) In either case a time series is not a brute fact but a way of displaying a collection of brute facts. Hence it is not an instance of an inherently unpredictable fact.

The case of research and invention is more difficult. We know that problems are in the habit of seeming to pop up unexpectedly, and that hypotheses and ideas for experiments or designs come unanticipated. And we also know that even the most carefully planned research is full of surprises, to the point that the initial problem may be replaced with another, and a new method may be invented to handle it. (It follows that no research plan should be detailed and rigid for, if it were, it would suffocate original research. Only routine research can be planned and forecast in detail.) It would seem, then, that originality is unpredictable.

Still, research leaders do make forecasts concerning the research they are responsible for. They make them on the basis of what is known now, as well as on their knowledge of the available human and material resources. Thus Crick (1966) forecast the synthesis of enzymes and even genes, which were actually performed a few years later; and Watson (1976), in his well-known textbook, lists a number of open problems and notes that many of them will probably be solved within the next decade. So, some forecasts of the outcome of original research *are* made successfully. (And, *pace* Popper, such predictions are not logically impossible, for they do not consist in solving problems but in hazarding that the problems will be solved.) But they are intuitive forecasts, not scientific ones: they extrapolate from current events. (Moreover, they are partly self-fulfilling forecasts, for they incite young researchers to take up certain problems.) A scientific forecast of scientific and technological creations would require an extensive and profound knowledge of the neural springs of creativity and of the social conditions that stimulate and stifle it. And we do not even know whether such knowledge will ever be attained. It would be just as rash to claim that it will be attained, as to state that it is beyond our reach. It is more reasonable and rewarding to work on these problems than trying to predict the outcome of such work.

5. CONCLUDING REMARKS

We all wish to understand what goes on in and around ourselves. Such understanding calls for answering what may be called *the six W's of science and technology: what* (or *how*), *where, when, whence, whither*, and *why*. In other words, understanding is brought about by description, subsumption, explanation, prediction, and retrodiction.

Neither of these epistemic operations is more important than the others. However, most workers specialize in only some of them. Thus collectors, classifiers and field workers are more interested in description than in explanation or prediction; on the other hand theoretical researchers and technologists want to know not only the facts but also why they are so, in particular what makes things tick. Subsumption, explanation, prediction and postdiction are typically theoretical operations, i.e. activities performed with the help of theories—the better the richer and the better organized.

Answering any of the six *W*'s of science and technology is intrinsically valuable: it yields new knowledge. But it may also be instrumentally valuable for allowing us to modify reality to our advantage. Thus if we know how a system works we may be able to keep it in good repair or even to perfect it. Explanation has therefore not only a cognitive but also a practical value. However, explanation is not enough: certain doctrines, such as fatalism, explain everything but, being predictively barren, cannot be put to the test and so cannot be evaluated on empirical grounds. Only theories that help predict or postdict can be contrasted with facts. However, the matter of checking deserves a separate chapter.

CHAPTER 11

PRODUCING EVIDENCE

If we proceed scientifically in any epistemic field, whether scientific, technological, or humanistic, we will try to check our hypotheses and recheck our data. We will not use authority, let alone ignorance, to justify our hypotheses or data. This maxim sounds obvious but it is not. To begin with, the argument from authority is still going strong in many humanistic fields—a remnant of the time when they were not distinguished from ideology. Even in science we cannot dispense with a modicum of authority if we are to use knowledge reaped by others: every time we borrow or quote a result obtained by other workers we hope our source to be "authoritative" (competent and responsible) and put our trust in it. However, this trust is not blind: it is the reasonable belief that the quoted author has proceeded scientifically, and we are ready to give it up the moment it is shown to be incorrect.

As for the argument from ignorance, it is of the form "Since we do not know exactly what X is, or how X came about, we can safely assume that X is Y". Assume, yes; safely, not, for every assumption must be checked. And yet this maxim is sometimes violated in science. Consider the following arguments from ignorance gleaned from the scientific literature. (*a*) "Since we do not know in many cases what the probabilities of the various alternatives are, we may assume that they are equal". (The correct policy is to assume equiprobability as the simplest though not necessarily the most likely hypothesis, and to proceed to check this hypothesis. In the absence of information all hypotheses are equally likely—or unlikely.) (*b*) "If an exception to a given well-tested law statement occurs, it must be for lack of precision in the observation or measurement." (Right policy: Redesign the observation or measurement to see whether the exceptions were genuine or "artifacts", or at least keep those exceptions in mind instead of dismissing them out of hand.) (*c*) "Since we cannot observe a microthing within an arbitrarily small time interval, we can assume that it does not obey any conservation laws between observations". (Wrong: the basic law statements do not say anything about conditions of observation, let alone non-observation. To suspend the laws of physics just because nobody is looking is shear solipsism—or insanity. Yet this is the usual justification for introducing the so-called "virtual" particles and processes. See Bunge, 1970.) (*d*) "Any apparent exception to biological adaptation must be just apparent: if we knew more we would realize that actually it is useful to the individual or the species concerned". (The right hypothesis is that many features of organisms are nonadaptive—e.g. because they are remnants of formerly adaptive traits. The presupposition that every property of an organism is adaptive clashes with the evidence of extinction.)

A characteristic of mathematics, science, technology, and the modern humanities, is that everything in them is checkable: every datum, theory, method, artifact, or plan of action is supposed to be able to pass some test or other, if possible multiple tests. To be sure testability, though necessary to attribute truth or efficiency, is insufficient to decide that an item of knowledge is scientific or technological. Thus astrology, though refuted long ago, never belonged to science; and alchemy, though shown to be both false and ineffective, was never part of technology. In addition to the possibility of testing we need actual testing and we cherish positive test results no less than negative ones. In particular we delight in learning that a theory has been confirmed or a machine actually does what it was designed to do.

We are interested in tests of two kinds: tests for truth and tests for efficiency. They concern objects of different kinds: the bearers of truth are propositions (or the underlying thoughts), and those of efficiency are actions and artifacts (or the underlying proposals, plans, designs, instructions, or norms). Thus whereas data and hypotheses can be true to some extent, norms and computer programs can only be effective to some extent. Truth guides the search for efficiency, and efficiency poses the problem of its explanation, but they are different categories. So much so that truth is often useless, and efficiency is sometimes based on untruth. Thus many true scientific hypotheses have no practical application, whereas some worthless products are commercially successful because of deceitful advertising.

Testability and truth, which are only accessory in the case of prescientific ideologies and public affairs, are central to science, technology, and the humanities. To be sure, doing research in either of these fields is investigating problems with the aim of solving them. But a problem is considered to be solved only if the proposed solution passes checks for truth (in the case of cognitive problems) or efficiency (in the case of practical problems). Therefore it is not true that the success of the scientific enterprise may be gauged by its ability to solve problems regardless of matters of testability and truth (as Laudan (1977) claims). A thoroughly untestable solution is not scientific or technological, and an untested one can be assigned neither a truth value nor an efficiency value.

1. FROM SELF-EVIDENCE TO EVIDENCE

1.1. *Self-evidence*

We are all more or less gullible in some department or other. In some cases we regard our beliefs as self-evident, in others we rely on authority, and in still others we refuse to admit unfavorable evidence. Let us make a quick study of these three cases.

Intuitionist philosophers of all hues agree that there are certain obvious or self-evident truths, i.e. propositions the truth of which is recognized intuitively without any tests, i.e. without evidence, and which serve as a standard relative to which all other propositions are evaluated. Husserl went to the extreme of claiming that some people are endowed with a vision of essences (*Wesensschau*) allowing them to grasp the essence of any object without having to waste their time experimenting or theorizing. But of course the history of mankind is littered with the corpses of supposedly self-evident truths, which are just received opinions and often blatantly false superstitions. What is obvious to the layman may be puzzling to the expert, and conversely. Besides, specialists do not behave homogeneously: what is surprising to one may be self-evident to another. But, if competent and responsible, they will not follow blindly their gut feelings but will test them.

Faith in intuition is not restricted to intuitionist philosophers. In all societies most laymen conform to the judgment, prognoses and decisions of experts, be they plumbers or electricians, physicians or economists, shamans or statesmen. We usually take their word although their judgment is more often intuitive rather than scientific—and often wrong to boot. In particular clinicians are known to propose diagnoses in a hurry and to fail in a high percentage of cases, and economists are prone to make recommendations on the strength of theories which, in the best of cases, have not been tested.

Blind faith, i.e. trust in ideas or procedures that have not passed sufficient tests, is not limited to the professions. Many hypotheses in social science have never been checked or, when tested, have turned out to be false—notwithstanding which they are often kept with ideological zeal. Here is a random sample of such views. (*a*) According to Marxism, the modern state is nothing but the supermanager of capitalism; in particular it always protects the monopolies and takes upon itself to run the low return and losing sectors of the economy. The profitability of the French and Canadian state corporations has conclusively refuted that thesis (Niosi, 1981). (*b*) According to neoclassical microeconomics prices are fixed by the play of offer and demand: the price of a good or service is that for which the offer equals the demand. Actually the big corporations fix prices without regard to demand (Galbraith, 1967; Silberston, 1970). (*c*) According to Durkheim, the rate of crime in any stable society is constant over time, so there is nothing one can do about it. The data, at least for California between 1851 and 1970, are inconsistent with this hypothesis and show that the crime rate fluctuates rather wildly (Berk *et al.*, 1981).

We all tend to protect our most cherished beliefs, particularly when we have invested much time and effort acquiring them—as is the case with theologians, political ideologists, and economists. Thus if someone believes that prayer (or vitamin C) will protect him from the common cold, he will pray (or take vitamin C) regularly; should he catch a cold in spite of this, he will probably double the dose rather than question his belief. Failure to admit negative evidence is a very natural disposition: skepticism is learned only the hard way. Jevons (1877, p. 402) remarked long ago that "It is difficult to find persons who can with perfect fairness register facts for or against their own peculiar view". Contemporary psychologists have confirmed this "law of the disregard of negative information". Thus psycholinguistic experiments have shown that "Most people prefer to accentuate the positive" (Miller, 1967, p. 168). Even if they get to know the worth of their false beliefs and wrong actions, most people make hardly any use of such feedback. (See Festinger *et al.* (1956), Wason and Johnson-Laird, (1972), and Shweder (1980), for some experimental material on the tenacity of belief in the face of contrary evidence.)

The most common attitude in the face of unfavorable evidence is then one of disbelief or indifference, or at most doubt. This attitude prevails not only in ordinary life but also in science: "facts that fit into a preconceived hypothesis attract attention, are singled out, and are remembered; facts that are contrary to it are disregarded, treated as 'exceptions', and forgotten" (Luria, 1975, p. 339). There are several possible and mutually compatible reasons for such conservatism. One is that negative evidence is seen as failure or even as threat or humiliation: discarding it or "rationalizing" it is then regarded—particularly in success-oriented communities—as necessary for the preservation of self-esteem. Another reason is sheer intellectual laziness: it may be too hard to modify one's body of beliefs to accommodate the fresh evidence. A third is social (economic, political, or cultural) conservatism tied to vested interests (economic, political, or cultural)—as is the case with the persistence of moth eaten beliefs such as economic laissez-fairism, political centralism, and religious fundamentalism.

However, conservatism is not all that bad: it all depends on what it attempts to keep. Besides, cognitive conservatism is deeply ingrained in our brain and, to a point, it is methodologically justifiable. Let me explain. First of all, psychologists know that—unlike computers—we do not accept new information the moment we acquire it, but screen it for compatibility with our belief system (Rokeach, 1960). If the new information is consistent with our belief system or irrelevant to it, we tend to accept it without much scrutiny; otherwise we tend to alter it to adapt to our fundamental beliefs—or we just reject it. In short, we evaluate new information as it comes in, attributing far more value to favorable than to unfavorable data.

Such epistemic conservatism is double-edged. On the one hand it blinds us to new unsettling information. On the other it spares us countless mistakes. After all, the data themselves may be incorrect: we did not look or measure carefully, the entire experimental set-up was vitiated by some systematic error (e.g. a leak), or we were hallucinating. So, it is methodologically sound to question and even discard data incompatible with a body of knowledge—provided the latter has been well tested. It is not that we must prevent ugly facts from spoiling our beautiful theories: we must take care not to accept all data at face value, and we cannot assign truth values except in some body of knowledge or other. (Imagine trying to evaluate a report on the spin of gluons without having the slightest clue as to the meaning of "spin" or "gluon".)

Although we are all natural epistemic conservatives, we can learn to check ideas and things, and to modify them or give them up if they fail utterly. Such learning is part of the training of every craftsman, manager, technologist, scientist, and modern humanist. Indeed all the sciences, formal or factual, basic or applied, as well as the crafts and technologies, and the modern humanities, are supposed to abide by the following *testability principle*:

RULE 11.1. Every datum, hypothesis, technique, plan, and artifact must be checked for adequacy (truth or efficiency).

In other words, no matter how a bit of knowledge has been acquired, how a design has been conceived, or how an artifact has been manufactured, they should pass certain tests before being accepted. (Reichenbach (1938) emphasized the difference between the "context of discovery" and the "context of justification". There was nothing wrong with this; his mistake was to believe that it was possible to construct an a priori logic of discovery of generalizations from observations, i.e. an inductive logic.) Whereas we should be permissive with regard to discovery and invention—within the bounds set by the public interest—there is little play when it comes to evaluating them. The checking must be rigorous, for we want truth or efficiency, as the case may be, and neither can be ascertained without tests.

This does not entail that, unlike checking, discovering and inventing are utterly lawless and irrational and therefore beyond the scope of scientific research. The history of science, technology and the humanities teaches us that new findings do not pop up out of the blue: that they answer questions posed before against a certain epistemic background, that they are acquired with the help of means evolved before, and that they lead rather naturally to new developments—provided no social hurdles are in the way. Besides, like all other cognitive processes, discovery and invention can be studied both as brain processes and as social processes. Uncovering their mechanisms and those of their neural and social inhibitions should help us discover and invent, or at least prevent us from interfering destructively with them.

To return to the testability principle (Rule 11.1). To be sure, it would be impossible for us to test every proposition, proposal, and artifact that comes our way: the life and resources of an individual are too short for that. Therefore checking is a social endeavor: we submit our findings to the examination of our peers, and trust most of the knowledge we borrow from fellow workers. But this trust is neither blind nor unshakeable: we trust only the information that has been screened, and we do so only provisionally, i.e. until shown wrong. In other words we adopt the *fallibilist* principle:

RULE 11.2. Regard every cognitive item—be it datum or hypothesis, technique or plan—as subject to revision, every check as recheckable, and every artifact as imperfect.

However, there are degrees of adequacy, hence of inadequacy: some propositions are truer than others, some methods more powerful or accurate than others, some plans more effective than others, some artifacts more suitable than others, and some checks more rigorous than others. Consequently radical skepticism—according to which everything is equally worthless—is as false as it is paralyzing. Only moderate (or methodological) skepticism may help us advance from one inadequate proposition to a less inadequate one, from one ineffective proposal to a less ineffective one, from one unsuitable artifact to a less unsuitable one. Such moderation is controlled by the *meliorist* principle:

AXIOM 11.1. Every cognitive item, every proposal, and every artifact worth being improved on can be perfected.

Of course there is no way of proving this optimistic principle, and this is why, if accepted, it must be adopted as a postulate. All we can affirm with confidence is that so far it has been confirmed—a case of self-fulfilling prophecy. However, the principle has its built-in limitation, namely the proviso that the item to be improved be worth being improved. There are two reasons for this proviso. One is that some items had better been dropped altogether—e.g. torture instruments, cigarette lighters, and leather covers for moustaches. The other is that even serviceable ideas, procedures and artifacts reach a point where any further investments in their improvement would by far outweight the returns on them.

To sum up, self-evidence is illusory, authority can (and must) be challenged, and negative evidence, though we naturally dislike it, lurks behind everything. Hence we must be fallibilists—but, at the same time, meliorists. Our "worlds" of knowledge, plans, and rational action are very far from being the best of all possible worlds, but they are not the worst either and, what is better, they are perfectible.

1.2. *Evidence*

If self-evidence is no evidence, what is? A theoretical consideration can be; also an empirical datum, although not all data qualify as evidence. We shall deal with conceptual evidence in the next chapter, Section 3.1. Suffice it to say here that every well-confirmed theory is evidence for any other theories presupposing the given theory—and vice versa. As for the condition a datum must satisfy to rate as evidence, it boils down to relevance. Let us elucidate these concepts.

A *datum* is a particular piece of information of either of the forms "Thing x is in state (or goes through process) y", and "There are things of kind (or possessing property) *K*". ("Thing *x* has property *P*" is a particular case of the former, since a state is a list of properties.) An *empirical datum* is a datum acquired with the help of empirical operations, such as observation, measurement, experiment, action, or a combination thereof. And a *genuine empirical datum* is an empirical datum obtained with the help of operations that are accessible to public scrutiny. This excludes revelation, hunch, and authority.

Not all data constitute evidence for or against an epistemic item, such as a hypothesis, or a value judgment concerning the efficiency of a plan or the performance of a machine. Only relevant data may constitute evidence. An empirical datum may be said to be *relevant* to an epistemic or an evaluational item only if it refers to the latter. If preferred, a datum is relevant to a proposition if both share at least one predicate. For example, information about the performance and price of a machine are relevant to any statement about the profitability of the machine. On the other hand price and performance are mutually irrelevant even though both refer to the same machine, for a machine may be expensive yet perform poorly, or inexpensive and perform satisfactorily. In other words, there are neither laws nor rules that relate price to performance in a regular way: this is why price and performance are mutually irrelevant. (They may be statistically correlated, though.) On the other hand data about crowding are relevant to hypotheses about social cohesion, because the latter happens to depend upon the former.

Relevance and possible lawful relatedness are not enough for a datum to constitute evidence for or against a proposition or a proposal. In addition, every datum must be interpreted as possible evidence. Such interpretation is built into the very construction of measuring instruments or it is explicitly laid out in their operating manuals on the basis of knowledge about their design and operation. Thus we have learned to read time from a watch, and energy from an electric meter. But in most cases data come uninterpreted: we do not know what they "mean" unless we are able to interpret them in the light of some body of knowledge, preferably a theory. For example, the lay may interpret reports on experiences of dying as evidence for life after death, whereas the psychologist will "read" them as hallucinations. And, whereas the traditional Chinese "read" certain fossil teeth as dragons' teeth, the palaeoanthropologist, armed with the theory of evolution, sees them as remains of ancient apes or hominids. In short, the raw datum is no evidence: it becomes evidence when suitably interpreted in line with some of our background knowledge. (cf. Bunge, 1967b.)

We summarize the preceding observations into the following definition. An empirical datum *e* constitutes *empirical evidence* for or against a proposition or a proposal *p* (another datum, a hypothesis, a plan, a value judgment, etc.) if, and only if, (*a*) *e* has been acquired with the help of empirical operations accessible to public scrutiny (rather than made up, conjectured, taken from authority, or obtained by allegedly paranormal means): (*b*) *e* and *p* share referents (or predicates); (*c*) *e* has been interpreted in the light of some body of knowledge, and (*d*) some regular association (law or rule) between the properties represented by predicates in *e* and in *p* is (rightly or wrongly) assumed to exist.

The empirical operations referred to in clause (*a*) above are seldom purely empirical, particularly in modern science and technology. Thus the land surveyor employs scientific instruments, such as theodolites, together with geometry. Physicists, chemists and biologists use far more sophisticated instruments in conjunction with even more refined hypotheses. And, increasingly, they also use computers to drive microscopes, time the taking of pictures, control micromanipulators or servomechanisms, order the taking of measurements, and plot and even process the results of the latter. In all these cases a number of complex theories are involved in the design, performance, and utilization of empirical operations. In short, refined and exact empirical data are anything but theory-free perceptual reports. More in Section 3.

Data are seldom error-free. In particular, quantitative experimental data are likely to be subject to errors of two kinds: systematic (deriving from bias or defective design) and random (deriving, e.g. from random thermal motion and random external perturbations). We shall return to this point in Section 3. Let us now note only that, because data may not be error-free, they must be checked instead of being taken at face value. The new (checking) run of empirical operations may be done with the same technique or, preferably, with a rival (equivalent or better) technique. (Thus the ages of terrestrial rocks are estimated with a variety of methods, such as radioactive decay, accumulation of helium produced by cosmic ray impact, and even models of the evolution of the solar system.) Furthermore, ideally the checking is done by independent workers in order to minimize personal bias. Calculations are parallel.

In sum, the results of observations, measurements or experiments are not to be accepted at their face value but must be interpreted and screened. In particular, numerical data must be corrected in a number of ways, e.g. for parallax, pressure dependence, and sampling—not to speak of bias or systematic errors. Once corrected, the data are subjected to statistical processing, which yields aggregate data such as averages together with random errors (or scatters). In the process some data get eliminated, either because they are outlying (hence likely to be effects of artifacts or bias), or because they are in flagrant contradiction with well-confirmed theories. (Some outstanding scientists, such as Ptolemy, Newton, Mendel, and Millikan, have been accused of fudging or doctoring data in order to save their pet theories. This accusation is somewhat unfair: every good scientist knows that no data are to be taken at their face value.) The net result of a run of empirical observations yielding numerical data is a table or schedule from which a graph can be constructed, which in turn suggests a function. In short, the entire process from raw data to final evidence is the following chain:

$$\textit{RAW DATA} \xrightarrow{\textit{Interpretation}} \textit{EVIDENCE} \xrightarrow{\textit{Correction}} \textit{CORRECTED EVIDENCE} \longrightarrow$$

$$\xrightarrow{\textit{Statstical processing}} \textit{AGGREGATED DATA} \longrightarrow \textit{PRESENTATION.}$$

Administrators are often reluctant to fund the checking of data and calculations: they tend to see it as wasteful duplication of research. Yet rechecking is of the essence of scientific research. Besides, strictly speaking there is never exact replication, for there are never two identical research teams working in exactly the same circumstances. If smart enough, one of the teams is bound to find something—a new fact, a new hypothesis, a new method, or a new problem—that the other missed. And the team engaged in checking the work of others is bound to find some errors in the assumptions, procedures, or results of the first team. The outcome may be a more thorough checking, perhaps to the point of demolishing the earlier work or even showing that it had been triggered by an ill conceived problem.

As a matter of fact checking and rechecking are not the prerogative of scientists and technologists. Every rational man checks before, while, and after he acts or thinks. The horseman (or motorist) checks his mount (or car) before mounting (or driving); the craftsman checks his tools and raw materials before working with them; the biologist checks his preparations and instruments before and after conducting his experiments; the theoretical scientist checks his assumptions and the results of his calculations. Checking and rechecking is of the essence of good craftsmanship and of the practice of science and technology, just as it is alien to pseudoscience and prescientific ideology.

Practical men must have appreciated the value of checking procedures and things for thousands of years, or else they would not have left descendants. But the realization of the value of checking ideas seems to have dawned only recently. The first proof of a mathematical statement seems to be only 2,500 years old, and the first scientific experiments were not performed until the 17th century. For thousands of years physicians have prescribed cures, judges passed sentences, and religionists pronounced dogmas, without sufficient evidence and sometimes with no evidence at all. Most of mankind has lived and continues to live affirming or denying, at best debating, ideas supposed to be so important that they were deemed to be above checking. Only the humble hunters or farmers, stonemasons or scribes, took it upon themselves to check some ideas about their own work while allowing their economic, political and cultural leaders to mislead, oppress and exploit them in the name of unchecked or false principles.

This situation has changed, at least in the sciences, technologies and some branches of the humanities. In these fields it is now recognized that all truth claims—such as "*p* is true" and "*p* is truer than *q*"—and all efficiency claims—such as "*p* is effective" and "*p* is more effective than *q*"—are to be justified or validated by some objective means or other. However, philosophers are still divided on the acceptable ways of justifying truth claims, and they have not reflected much on the ways of validating efficiency claims. Leaving aside intuitionists—for they care only for self-evidence—and conventionalists—for not being interested in truth—we recognize the following philosophical camps with regard to the problem of the justification or validation of truth claims: see Table 11.1.

Unanimism (e.g. Ziman, 1979) holds that intersubjective agreement is necessary and sufficient for validation. Counterexamples: (*a*) the agreement of a thousand theologians concerning the existence of a deity does not prove it; (*b*) radically new theories are likely to be initially rejected or ignored by the great majority. *Pragmatism* (e.g. James, 1907) equates truth with success and therefore accounts for neither. We all know that some successful doctrines are false, whereas some true theories have found no practical application. *Rationalism* (e.g. Leibniz, 1703) holds that ideas are tested by checking their coherence with a body of basic (and possibly innate) beliefs. But coherence is insufficient: think of any consistent ideology and of that of some madmen. Besides, there is no reason to exempt

TABLE 11.1
How different philosophies of knowledge view the validation of truth claims

Philosophy	*Necessary and sufficient condition for validation*
Unanimism	1 Consensus of experts
Pragmatism	2 Practice (individual or social)
Rationalism	3 Coherence with background knowledge
Empiricism	4 Positive evidence
Critical rationalism	5 Lack of negative evidence
Critical realism	6 Numbers 3 through to 5 above

the principles themselves from rigorous checking. Nevertheless rationalism has an important grain of truth, namely that coherence is necessary: we dismiss hypotheses that are blatantly inconsistent with data, as well as data that fly in the face of well-confirmed theories.

Empiricism (e.g. Reichenbach, 1938) too contains an important grain of truth: experience is indeed a (though not the) test of truth and efficiency. In particular, positive evidence for the predictions calculated with the help of a theory confirm the latter. *Critical rationalism* (e.g. Popper, 1959) agrees that experience is a test of theories (its only concern) but claims that only negative evidence counts (against), for positive evidence is too easy to come by. True, unsuccessful attempts to refute a theory (or discredit a proposal or an artifact) are more valuable than mere empirical confirmation. However, (*a*) the most general theories are not refutable, although they are indirectly confirmable by turning them into specific theories upon adjoining them specific hypotheses (Bunge, 1973b); (*b*) true (or approximately true) predictions are not that cheap, as shown by the predictive barrenness of pseudoscience; (*c*) positive evidence for the truth of an idea or the efficiency of a proposal, procedure, or artifact, does count: thus the US Food and Drug Administration will rightly demand positive evidence for the efficiency of a drug before permitting its marketing.

Critical realism is a sort of synthesis of rationalism (the coherence requirement), empiricism (positive evidence), and critical rationalism (negative evidence), plus the realist thesis that the theories in science and technology represent (poorly or accurately) parts or aspects of the real world. I submit that critical realism is the tacit epistemology and methodology of science and technology. True, there is never overall coherence because new ideas are bound to be inconsistent with some bits of background knowledge; but this only shows that the latter is changeable and that justification or validation is never final. True, all experience is partial and inconclusive; but it can be enriched and checked by further experience as well as by theory. True, lack of negative evidence is valuable; but it is not enough, particularly since it is not more certain than positive evidence. In sum, critical realism demands at the same time coherence with the (changing) background knowledge, substantial positive evidence, and lack of significant negative evidence. It demands it not only from science and technology but also from the nonformal branches of philosophy, namely ontology, epistemology, and ethics. This empirical constraint on philosophy allows one to write off, as mere phantasies, such doctrines as that the mind is immaterial, that all knowledge is a matter of convention, and that moral behavior is unaffected by society. We need not waste much time in examining the ingenious arguments offered to buttress such phantasies: what matters is that they do not agree with the facts—as shown by their incoherence with our background knowledge and the absence of any positive evidence for them. (More on critical realism in Ch. 15, Sections 2 and 3.)

Note that we have been concerned with truth and efficiency claims rather than with beliefs. (Recall Ch. 2, Section 3.2 on the difference between knowledge and belief.) On the other hand the currently fashionable epistemologies concern belief. In particular the so-called coherence theories (which are really views rather than theories) hold that a belief is justified only if it coheres with all the other beliefs in a given belief system; and the "foundation" theories add that those beliefs that are most certain are to be taken as the foundation of our knowledge. Such epistemologies, which constitute a version of traditional rationalism, are ideally suited to religious and political ideologies, for coherence is the most that such bodies of belief can claim. By the same token those epistemologies are unfit for science and technology: here we are supposed to know (however provisionally) that a hypothesis has a high degree of truth, a procedure a high degree of efficiency, or a machine a high degree of reliability, before we believe in their adequacy. And such knowledge must be an outcome of research, both empirical and theoretical, not just a matter of checking coherence with a body of beliefs.

We hold that the study of belief, though a legitimate concern of epistemology, must be conducted like the study of any other real process, namely empirically as well as theoretically. Belief is mainly a concern of individual and social psychology. An armchair study of belief, yielding an a priori theory, is necessarily poor and dogmatic: poor because it makes no use of experiment, and dogmatic because it does not care for empirical validation. Take for instance the traditional beliefs on the correlation between ideology and social class. No a priori speculation is an adequate substitute for an empirical investigation into the actual correlation between, say, belonging to the working class and sympathizing with socialism, or belonging to the petty bourgeoisie and voting liberal. (See, e.g. Boudon, 1967.) Another example: only empirical investigations have established that (*a*) many popular superstitions (e.g. belief in telepathy) are formed by the natural tendency to conjecture causal relations, even about conjunctions of events that are nothing but coincidences or statistical flukes (i.e. chance accumulation or thinning of events in random sequences), and (*b*) people who are properly trained in probability theory and statistics are less prone to forming such beliefs. (See, e.g. Tverksy and Kahneman, 1977; Falk, 1982.) In sum, only scientific research into belief can yield genuine knowledge of belief: the nonscientific study of belief yields only further belief. And only scientific or technological research can tell us under what conditions it is reasonable to hold, withhold, or reject specific beliefs. In short, first knowledge through research—in particular knowledge about truth and efficiency—then credence.

In conclusion, claims of truth or efficiency must be validated by some evidence. Evidence can be conceptual or empirical. (However, the support a factual theory lends a proposition or a proposal may be regarded as indirect empirical evidence. And an empirical datum is evidence for or against a proposition or a proposal just because some theory suggests so.) Evidence can be positive, negative, or inconclusive. And in principle neither positive nor negative evidence is definitive, if only because further research may show that it is inadequate.

2. TESTABILITY AND INDICATORS

2.1. *Testability*

A proposition will be said to be *empirically confirmable* if there is direct or indirect, actual or potential empirical evidence for it; and *empirically refutable* if there is direct or indirect, actual or potential empirical evidence against it. A proposition that is only confirmable, or only refutable by empirical data, will be said to be *testable*. Finally, a proposition will be said to be *strongly testable* if it is both confirmable and refutable, and *untestable* if it is neither. (Concepts are testable for relevance and power but not for truth: only propositions are testable for truth.)

Because empirical data are obtainable by certain means but not others, and only provided the state of the art has attained a certain level, testability is not an inherent property of propositions but is *relative* to the available or conceivable empirical means. It may happen that a proposition is testable by data of a certain kind but untestable by data of another; or that it is better testable by data of one kind rather than another; or that it is more testable than another proposition relative to a given body of data. In sum, testability comes in degrees and is relative to the (empirical and conceptual) test means. So, any theory of degrees of testability would have to elucidate expressions of the forms "The testability of p relative to means m equals t", "p is better testable by m than by m'", and "p is more testable than q with the help of m".

Note that we have made room for potential data or, what amounts to the same, for testability *in principle* alongside actual testability. For example, a historical hypothesis conceived to explain certain data (or their absence) is likely to suggest the search for new evidence of a certain kind, which search may stimulate the invention of new techniques or the application of techniques evolved in other research fields. Another example: it is not always possible to determine, with today's means, whether a certain enzyme is present in a given chemical system or organism, because the substrate-enzyme complexes are unstable and therefore short lived. So, many conjectures in biochemistry, though potentially testable, are not yet actually testable. (According to the well-confirmed Michaelis–Menten hypothesis, a reaction of the form $A \rightleftharpoons B$ catalyzed by C actually summarizes two reactions involving the substrate-catalyzer complex AC. They are: $A + C \rightleftharpoons AC$, which is reversible, and $AC \rightarrow B + C$, irreversible. If the corresponding rate equations fail, we hypothesize that a different or a further catalyzer is at work.)

We have also included *indirect* empirical evidence, i.e. evidence through some intermediary body of knowledge. For example, "There have been dinosaurs" is not directly testable, for there are no living dinosaurs nowadays but only certain fossil bones, eggs, and footprints. The hypothesis itself allows one to interpret such data as evidence for it. (Other historical hypotheses are parallel.) However, far from being a stray conjecture, the dinosaur hypothesis is part of the vast body of evolutionary biology, which in turn is consistent with the rest of biology as well as with geology. Or take the hypothesis of interactionist mind-body dualism, namely "Mind and body are distinct though mutually interacting substances" (Cf. Popper and Eccles, 1977.) This hypothesis is confirmable because it accommodates any conceivable mental event; by the same token it is not directly refutable. However, the hypothesis is indirectly refutable, namely by showing that it is incompatible with physiological psychology and even with physics. Indeed any action of an immaterial mind on matter (e.g. a brain) would violate the principle of conservation of energy. (Cf. Bunge, 1980a.)

Other hypotheses are confirmable but irrefutable in principle, hence only weakly testable. For example, the hypothesis of hypnotic trance is immune to refutation because, if the subject obeys instructions, he is declared to be in such a state, whereas if he does not, he is said not to have been properly hypnotized (Barber, 1970). Another hypothesis of the same methodological type is "All persons seek to maximize their utilities". This axiom of utility theory (and of much of neoclassical economics) is confirmed, never refuted, by whatever any person does: even altruism and suicide can be interpreted as maximizing the person's moral if not economic utilities. Therefore it is weakly testable.

Finally let us exemplify the notion of an untestable proposition. Utter untestability can be due to either imprecision or the positing of inscrutables. Gestalt psychology is an example of a system of propositions that, because of their vagueness, are hardly testable (Claparède, 1934, p. 145). In such cases one does not know exactly what is to be tested. However, some imprecise propositions are worth being exactified and rendered testable. (*Example*: "Curiosity can be smothered or trained".) Note that exactness is necessary but not sufficient for testability. Indeed any proposition positing the existence of inscrutable entities is untestable even if precise. Here is a random sample of untestable propositions of this kind. (*a*) "We are surrounded by things that are unknowable in principle". (Kant's thing in itself was of this kind for him.) (*b*) "Physical space is embedded in a higher dimensional space, but we happen not to have access to the extra dimensions". (*c*) "There are things deprived of energy". (Such things could not possibly exchange energy with ordinary things, such as particles or fields, so they would go undetected.) (*d*) "Good people go to heaven, bad ones to hell".

There can be no empirical evidence for or against vague propositions or for propositions positing inscrutables. The same holds for some counterfactual conjectures, such as "If Einstein had not invented general relativity then someone else would have done it". All such propositions are empirically untestable. However, they may be discarded by the simple philosophical manoeuvre of asking their propounders "How do you know?" or "What evidence have you got?" (Parenthetically this mode of refutation is not available to Popper (1963), who regards such questions as inappropriate and, in general, matters concerning our information sources as unimportant.)

Let us now examine certain kinds of hypotheses whose methodological status seems uncertain: partial differential equations, ad hoc, existence, continuity, possibility and probability hypotheses, and estimates of nuclear weaponry capability. *Partial differential equations*, a standard tool of physicists, astronomers, engineers and many others, have infinitely many solutions. This richness is theoretically valuable but methodologically embarrassing, for it renders such equations testable in too many ways. (*Example*: any differentiable function of the argument $x - vt$, where x and t are independent variables and v a constant, solves the equation $(\partial u/\partial t) + v(\partial u/\partial x) = 0$. The latter may be partially interpreted as "For every entity k of species K, the u-ness that k gains or loses in the course of time is compensated for by a loss or gain in the u-ness of k effected by a change in place". Among others, field-like and particle-like solutions of the equation exist. For a detailed analysis of a theory generated by that hypothesis see Bunge (1967d).)

In other words, a partial differential equation offers a large target to empirical tests, but the results of these are inconclusive for bearing on the solutions not on the equation itself. (The situation is even worse with regard to the Lagrangian functions that, via extremum principles, generate those equations. Indeed every equation of motion or field equation can be generated by any of an infinity of such functions.) In other words, not the equations themselves together with their infinitely many consequences, but only some of the latter are confirmable by empirical data: see Figure 11.1. If data of one kind fail to confirm a given solution, then data of another may confirm it. Likewise if data of a given kind refute one of the solutions, they may confirm another. So, the original equation itself can hardly be said to be refutable—at least not until all its solutions have been tested. However, in practice the decision is considerably eased by three considerations. One is that a partial differential equation is usually one component of a whole problem system that contains also a number of restrictions such as initial or boundary values, which restrictions narrow down considerably the set of possible solutions. Another is that not every solution is interpretable in suitable factual terms; those which are not so interpretable are just

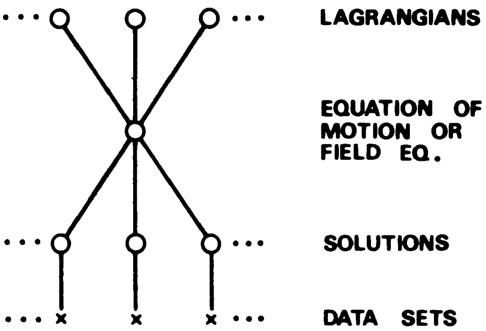


Fig. 11.1. One and the same equation of motion (or field equation) is derivable from alternative Lagrangians, and has different solutions. Data of a certain kind may favor one such solution.

dropped. A third is that the equation entails not only its own solutions but also a number of additional theorems—perhaps conservation laws—some of which are empirically testable. Nevertheless the evaluation of such high level hypotheses is anything but straightforward.

Next in our list of hypotheses that arouse some methodological unease are *ad hoc* hypotheses. An *ad hoc* hypothesis is a particular assumption, i.e. one covering a rather narrow range of facts. We distinguished two kinds of *ad hoc* hypotheses: *bona fide* and *mala fide*. (Ch. 8, Section 4.2.) A *bona fide ad hoc* hypothesis is normally offered to represent facts of a certain kind and it is independently testable. For example, to complete his theory of the circulation of the blood Harvey had to conjecture the existence of tiny vessels, that remained invisible during his lifetime, connecting arteries and veins; this hypothesis was confirmed later on with the help of the newly invented microscope. On the other hand a *mala fide ad hoc* hypothesis is designed exclusively to protect another hypothesis, or a theory, from refutation: it is a cyst. A classical example is Freud's repression hypothesis, which saves the Oedipus complex fantasy in case the individual fails to exhibit such "complex". Another is the parapsychologist's claim that the presence of skeptics inhibits the mind reader and the clairvoyant. *Mala fide ad hoc* hypotheses are to be avoided because they block the test process. On the other hand there is nothing wrong with *bona fide ad hoc* hypotheses. The earth sciences teem with testable hypotheses that, presumably, hold only for our planet, and are therefore both *ad hoc* and *bona fide*.

The *existential hypotheses*, i.e. conjectures of the form "There are *F*'s", are logically modest but ontologically ambitious and therefore methodologically tricky. Indeed they can be confirmed by exhibiting instances but they are hard to refute for, if an effort to find *F*'s fails, one may still hope that more strenuous efforts will meet with success. But of course no such efforts will be made unless there is some evidence, empirical or theoretical, for the existence of *F*'s. Columbus had both when he undertook his first voyage; so had Le Verrier and Adams when they predicted the existence of Neptune, and Hertz when he predicted the existence of electromagnetic waves.

Mere argument neither proves nor disproves existence—except of course in theology. In formal science existence is either postulated or proved from postulates with the help of logic. In factual science and technology existence conjectures are supported or undermined by both theory and empirical evidence. The former suggests, the latter establishes more or less conclusively. And in a science-oriented philosophy existence claims should be treated in like manner: they should be postulated or proved rigorously in the case of constructs, and supported by empirical evidence in case they concern concrete things. No matter how seductive an argument for the existence of deities, disembodied minds, propositions in themselves, and the like, it will cut no ice in scientific philosophy unless it can be supported by empirical evidence.

To be sure, theoretical considerations in science or technology, though powerless to prove or disprove the existence of any concrete thing, can render an existence hypothesis likely or unlikely, hence worthy or unworthy of being investigated. This they can do by showing that the given existence hypothesis is necessary for consistency or, on the contrary, leads to contradiction—as was the case with the aether hypothesis. Or such theoretical considerations may show that, if the entity in question were to exist (or not to exist), certain well-confirmed theories would fail. Thus no physicist can admit the existence of an entity that fails to conserve energy, or the nonexistence of a massive body, even if invisible, at a focus of the elliptical orbit of a celestial body. Still, the most conclusive proof of existence, whether in mathematics, science, or technology, is to exhibit a specimen of the object whose existence has been postulated.

Being empirically irrefutable, existence hypotheses have been declared untestable and therefore unscientific or metaphysical (Popper, 1959; Lakatos, 1978). However, this only goes to prove that testability is not to be equated with refutability. First, if we do find an *F* we have proved (until new notice) that there are *F*'s. (To be sure such evidence is not final—but then no empirical evidence is.) Second, the refusal to admit existential hypotheses entails giving up some of the most powerful and best confirmed scientific theories. Take for instance the core of classical mechanics (or of its relativistic generalization): "For all bodies and all forces and stresses, at all places and all times, *there exist* reference frames relative to which the rate of change of the linear momentum (or the momentum density) equals the body force plus the divergence of the stress tensor". Finding one reference frame relative to which this equation of motion holds, confirms the hypothesis; such a frame is called *inertial*. And finding any number of frames relative to which the hypothesis fails (e.g. our own planet) serves only to indict the frames as being non-inertial, not the hypothesis as false. Third and most important, the refusal to countenance existence hypotheses just because they are irrefutable, though consistent with conventionalism and pragmatism, is inconsistent with any realistic epistemology, for which the aim of science is to understand existents (and real possibles). This epistemology is inherent in the careful formulation of any scientific theory, where one begins by assuming that the reference class of the theory is nonempty—i.e. that the entities described by the theory *exist* on their own.

Still, could it not be that a theory passes some empirical tests and yet some or even all of its referents do not actually exist? After all there are plenty of historical examples of theories postulating nonexisting objects, such as caloric, aether, and magnetic monopoles. A first answer is that no final certainty is attainable in this case or in any other: the most one can obtain is solid though corrigible evidence. Second, a realist epistemology, one regarding the postulated entities as possibly existing rather than being fictions, will be more helpful than either an idealist or a conventionalist one. For, even if the objects postulated by the theory prove to be ghostly, the search for them is likely to yield some knowledge. (For example, the failure to find magnetic monopoles confirms once again standard electrodynamics.) Thirdly, there are no separate and fool-proof existence tests. In short, existential hypotheses are at the very core of science although, being only confirmable but irrefutable, they are not strongly testable.

Continuity hypotheses are of the form "The function F, which represents the property P of things of kind K, is continuous (or piece-wise continuous) over the domain D". Obviously, no measurement could possibly provide a direct test of such a hypothesis, for only a finite number of values of a function can be measured. So, we must resort to indirect empirical tests, such as those bearing on some consequence of the hypothesis. An example will elucidate this point. Suppose the available data are consistent with the following alternative hypotheses concerning a certain force: (*a*) the force varies continuously with the inverse of the distance, and (*b*) the force varies with the inverse of the distance when the latter is a rational number but is zero otherwise. To decide between the two hypotheses we may measure the work done by the force along a certain stretch. The result will be a nonzero quantity in the first case and zero in the second. (The work equals the definite integral of the force over the given distance. The integral is null in the case of the second function because it is defined on a set of zero measure.) In sum, continuity hypotheses are empirically testable though in an indirect manner. Moreover they are not only confirmable but also refutable: indeed any significant discontinuity in the corresponding property of their referents will refute them.

A *possibility hypothesis* is, of course, one stating that facts of a certain kind are possible. (In turn, real possibility equals compatibility with law: see Vol. 3, Ch. 4.) In science and technology possibility hypotheses are stated on theoretical considerations. For example, a spectral line or band is declared possible if the corresponding atomic or molecular transition is compatible with the atomic or molecular laws, and "forbidden" otherwise. And a technological project is declared to be in principle feasible if it involves no violation of any known laws—otherwise in principle unfeasible. Clearly, possibility hypotheses are confirmable but irrefutable on purely empirical grounds. The only possible empirical evidence relevant to a possibility hypothesis is of the positive kind: we prove that X is possible the moment we see X happen either naturally or through human intervention. (If X fails to happen we can still argue that it is too soon to write it off—or that, given the changed circumstances, it is no longer possible.) Actuality is then the only empirical evidence for possibility. (But this does not mean that actuality is the meaning of possibility, e.g. that probabilities mean frequencies.)

As for *probabilistic hypotheses*, they all pose the following problem. If the set of events under observation (i.e. the sample space) is smallish, whatever happens may be regarded as confirming the hypothesis; and if nothing happens the hypothesis cannot be said to have been refuted, for accidental crowdings and thinnings are of the essence of chance. For this reason a number of authors, among them Popper (1959), have regarded probability hypotheses as untestable or nearly so. However, this difficulty disappears as soon as huge masses of events become available: in such cases the probability of any deviation between probability and observed relative frequency approaches zero. This is actually the case with the various probabilistic hypotheses involved in particle and field physics, statistical mechanics, quantum chemistry, genetics, and to some extent social science as well. Thus even if the probability of an individual atomic collision (or genic mutation, or conversion of proletarian into tycoon) is extremely small, collisions (or mutations or dramatic changes in economic power) will happen in large aggregates of individuals or in the very long run.

Randomness is characterized not only by stable (or regularly varying) long run frequencies, but also by short term fluctuations, i.e. occasional high or low scores—e.g. streaks of good or bad luck. Only if the scores are consistently higher or lower than the calculated probability can we infer that the latter has been refuted. Typically parapsychologists ignore this peculiarity of random events when claiming that the psychic powers of their sensitives is "above chance" during certain periods and then—particularly when watched by skeptics—decline. In short, probabilistic (or stochastic) hypotheses are empirically testable by huge masses of like events: a single unfavorable case will not refute them. However, this only shows that we must make a philosophical decision: either we give up some of the deepest and best confirmed scientific theories—such as the quantum theories and genetics—or we reject any methodology that regards probabilistic hypotheses as untestable, and therefore unscientific, just because they are not tested the same way as nonprobabilistic hypotheses.

Let us now discuss the testability of hypotheses concerning the efficiency of weapons or strategies of certain kinds. As is well known, the Nazis subjected Britain to an intense bombing—only to unite and strengthen the Britons as well as to elicit universal sympathy and solidarity with them. Ignoring this teaching, subsequently the U.S. and British air force subjected Germany to an intense bombing in order to destroy the German economy and, in particular, its arms industry. This operation went on for nearly three years without the slightest indication of its effectiveness. Shortly after the war the U.S. Strategic Bombing Survey found that the bombing had been a dismal failure or worse, namely counterproductive. In fact the German arms production increased under the bombing: the production of tanks tripled, and that of planes doubled between 1942 and 1945. (Apparently the military did not relish this finding that refuted one of its most cherished dogmas: see Galbraith, 1981, Ch. 13.) Nowadays we face an even tougher problem. For the first time in history man has produced engines, namely intercontinental missiles with nuclear warheads, that are untestable. Indeed, if any of them were put to the test World War III could start, and presumably there would be nobody left to evaluate the damages. So, nobody can know exactly what those engines are worth—except of course in terms of the impoverishment of our lives. In this case speculating over the effectiveness of such artifacts is preferable to testing them. And in any event it is ironic that technology, with the help of testable and moreover well-confirmed theories, should have produced untestable artifacts.

Let us now compare constructs of different types as to testability. Obviously, such comparison is greatly facilitated if the constructs represent the same aspect of things of the same class, such as radiating antennas. Consider on the other hand the equivalence "For every x, if x is an organism, then: x has genotypic trait G iff x has phenotypic trait P". Since phenotypes are more easily recognizable than genotypes, the LHS of this equivalence is harder to test than its RHS. Equivalents are not necessarily equitestable. (*Methodological moral*: Logic is a poor guide to investigate testability.)

Assuming unlimited empirical means, singular propositions are better testable than universal ones, which in turn are more fully testable than hypothetico-deductive systems (theories). Thus "That antenna is radiating electromagnetic waves of such and such wavelengths in such and such direction" can be checked more fully than any particular formula of the theory of antennas, which theory is in turn even harder to test. Not that testing a singular proposition, i.e. one concerning an individual, is always plain sailing. Think of the extraordinary difficulties met in testing singular propositions of the form "That astronomical object is a black hole", and "This organism has just undergone a mutation". In neither case do observations suffice: in both cases the tests are extremely indirect and their results rather inconclusive.

General propositions are even harder to test than the corresponding singular propositions if the former concern a large or variable collection of individuals, as is the case with any basic law statement. True, in principle a single unfavorable case refutes the claim to universality—provided the evidence is reliable. But such result does not dispose of the generalization. In fact, though false in a given range, the generalization may hold in another, and this can be established only by continuing the tests after the first disconfirmations are in. Never mind the real motives of the scientist or technologist performing such tests: he may wish to confirm or to disconfirm. What matters is that the generalization can be pronounced false if nearly all the tests, over the whole range, are negative, and true (to some extent) in a given range if nearly all the tests in that range turn out to be positive. However, it may also happen that the tests are inconclusive, in which case a new, more precise testing technique may have to be tried.

Paradoxical as it may sound, single generalizations are in general harder to test than generalizations included in theories. The reason is that, whereas an isolated generalization can count only on the evidence bearing directly on it, a hypothesis belonging to a hypothetico-deductive system can also count on whatever evidence favors other components of the system. In other words, a systemic hypothesis can enjoy both direct and indirect empirical support. See Figure 11.2. Such indirect support is multiplied if the theory in question agrees with other theories in the same or different fields. For example, whereas a biological (neurophysiological) theory of mental functions can be supported by neuroscience and, indeed, by the whole of biology, a dualistic theory of mind cannot enjoy such support.

A theory is composed of infinitely many propositions (Ch. 9, Section 1.2). Therefore it can never be exhaustively tested: one must always confine oneself to testing a finite subset of the infinite set of propositions. Now, scientists do not test arbitrarily selected theoretical propositions but only those that interest them most—e.g. because they are the most novel—or that lend themselves more easily to tests. This tends to produce clusters of frequently tested propositions, alongside with pockets of propositions that are seldom if ever checked. There would be nothing wrong with this procedure if all the propositions of a theory had the same truth value, but this is not the case. Even highly successful theories, such as quantum electrodynamics, contain false propositions alongside others which are true or nearly so.

A solution to this problem would be to treat theories the way concrete populations are handled, namely to draw random samples. We might collect all the known and testable propositions of a theory, assign them random numbers, and draw lots. In this way the theory could be likened to

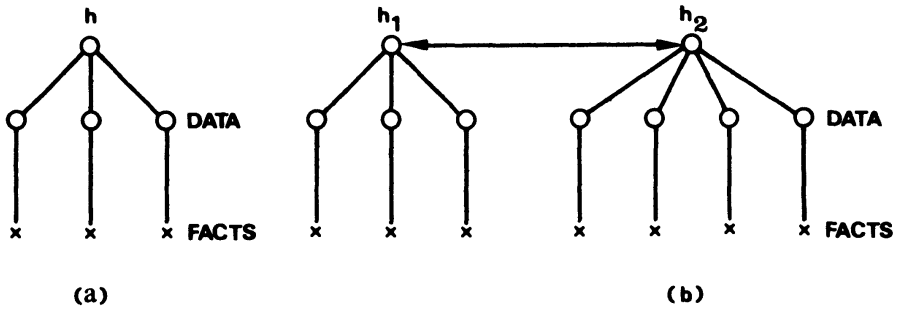


Fig. 11.2. The empirical support of hypotheses. (*a*) A single hypothesis is supported or undermined only by the facts it refers to. (*b*) A hypothesis belonging to a hypothetico-deductive system (theory) is controlled not only by direct evidence but also by whatever data are relevant to the hypotheses to which it is logically related.

a bag containing infinitely many marbles of unknown colors in unknown proportions. In the simplest case there would be only two colors, say white (T) and black (F), in unknown proportions. A test of the theory could thus be likened to the extraction and examination of marbles, one at a time and without replacements—except of course when one has reason to suspect some data and thus repeats the test. What are we to conclude if 100, 10,000 or even 1,000,000 balls turn out to be all white (T) or all black (F)? Strictly speaking *nothing*, for (a) there are still infinitely many unchecked marbles in the bag, all of which could prove to be black (F) in the first case and white (T) in the second; and (b) although the tested sample is a random one, the original population—in turn a finite subset of the set of all the propositions in the theory—was not, for we can know only those propositions in the theory that, for some reason or other, happen to have attracted the attention of theorists. And if the outcome of our test is that w marbles out of a total of n turned out to be white (T) and b black (F), all we can say is that the degree of confirmation of the theory (up to the nth test and with the given means) is $w/(w+b)$. But, for the above reasons, this datum is not very helpful except when the total number of tests is very large and the degree of confirmation is either nearly 1 or nearly 0. In the first case we may conjecture that the theory is true, in the second that it is false. But of course we must keep an open mind, as a further run of tests may force us to change our evaluation. Truth values, like the value of money, tend to depreciate in time.

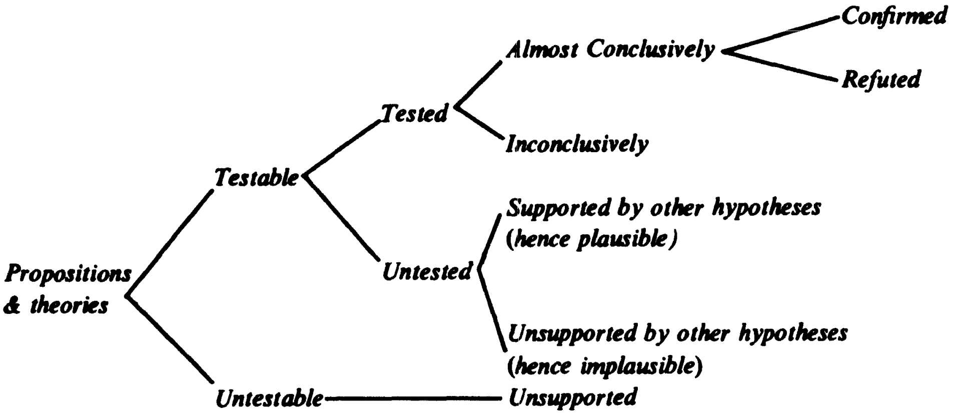
If theories in general are hard to test, the most general among them are the hardest. To be sure, if testable at all they are confirmable, but not necessarily refutable as well. Take for instance the most basic formulas of classical (or relativistic or quantum) mechanics. Undoubtedly they do have some general testable consequences such as the conservation theorems. However, in order to test any basic hypotheses we must adjoin them specific assumptions concerning the precise composition, environment, and structure of the system in question, be it atom, binary star, or machine, for we will contrast the theory with data concerning objects of certain kinds, such as atoms, binary stars, or machines. That is, we must conceive of a model object couched in the language of the theory: one specifying, say, what the components of the system are and how they interact with one another and with the environment. (Shorter: we must specify the Hamiltonian or the forces and stresses, the constraints and the initial and boundary conditions, and the constitutive relations characterizing the type of matter in question.) The outcome of this theoretical activity is a model, or specific theory, representing the object in question. This, not the general theory to which it is bound, is put to the test (Bunge, 1973a).

Since models (or specific theories), not general theories, are the ones that are subjected to empirical tests, the outcome of the latter, if negative, does not tell us unambiguously what went wrong: whether the general theory or the special assumptions adjoined to it. To be sure, uncertainty shrinks if a variety of models using the same general theory either succeed or fail. But even so some uncertainty remains. The only clear cut cases are exceptions: when none of many models grafted on to a general theory is confirmed, we are justified in blaming the latter; whereas if only a few fail, we blame the special assumptions. This methodological predicament is common to all extremely general theories. Thus the theory of evolution is so general as to be confirmable but not refutable. It becomes also refutable, and thus strongly testable, only when the genetic composition and the environment are specified, i.e. when applied to a particular lineage, over a long period of time. (See Bunge, 1978b.) What holds for the extremely general scientific theories holds, a fortiori, for the philosophical ones that concern the real world or our knowledge of it. Only the specific scientific theories based on those philosophical theories are strongly testable (Bunge, 1973b).

Let us remember that every factual theory, whether scientific, technological, or philosophical, has two ingredients, namely the formalism and the interpretation. So either or both can go wrong, and in some cases a correct interpretation can somehow compensate for a defective formula or conversely. Therefore if the theory is disconfirmed we may try altering portions of its formalism, its interpretation, or both. And although this is an eminently theoretical job, it cannot be conducted except in the light of the available empirical evidence. For example, there is consensus that the mathematical formalism of quantum mechanics is essentially correct, but there is disagreement about the correct physical interpretation of the theory. Yet it is possible to choose among the existing rival interpretations on both theoretical and empirical grounds. In particular, we can refute the subjectivistic interpretation of quantum mechanics just by noting that, if it were correct, (*a*) the theory would include psychological concepts (such as those of the state of knowledge and the intentions of the experimenter)—which it does not; consequently (*b*) psychological data (about the observer) would be relevant to both calculations and measurements—which they are not; and (*c*) the outcome of the empirical tests would depend critically on who performs them—which it does not. (More in Bunge (1967c) and (1973a).)

Finally, what about metaphysics or ontology? If defined as an untestable attempt to account for the world, the problem of the testability of metaphysics does not even arise. In this case metaphysics may be granted some heuristic value but it must be denied any truth value. To be sure, most of traditional metaphysics is untestable when intelligible at all. However, metaphysics can be overhauled with the help of exact tools and in such a manner that it becomes continuous with science: see Volumes 3 and 4. For example, it is possible to build a general and exact theory of spacetime compatible with contemporary physics, and a general and exact theory of mind harmonizing with contemporary physiological psychology. Therefore the initial definition must be altered to read: Metaphysics (ontology) is a family of (indirectly) testable theories about the world.

We close this section with a classification of factual propositions and systems of such with regard to their testability.


Propositions & theories
Testable
Tested
Almost Conclusively
Confirmed
Refuted
Inconclusively
Untested
Supported by other hypotheses (hence plausible)
Unsupported by other hypotheses (hence implausible)
Untestable
Unsupported

2.2. *Indicators*

According to radical empiricism scientific theories are data summaries, i.e. they have an empirical (not just factual) content. If this were true, then the testing of any such theory would be a straightforward contrasting of propositions of the theory with the corresponding data. But that is not true: scientific theories have a *factual* reference but, unless they actually describe human experience, they have no *empirical* content.

That scientific theories are not just data summaries, or even mere devices for churning possible data (e.g. predictions), is shown by the fact that they do not refer to the observations, measurements, or experiments aiming at testing them. Thus theoretical mechanics does not describe dynamometers (although it helps design and explain them), and genetics does not describe electron microscopes, which are employed in imaging chromosomes. Instead, scientific theories contain theoretical concepts lacking a counterpart in sense experience, such as those of continuity, free motion, strain, electric field intensity, valence, DNA, neural connectivity, GNP, and theory. As a consequence scientific theories contain propositions without an empirical counterpart. A simple example is that of the "principle" of inertia, a cornerstone of physics and a theorem of classical and quantum mechanics. The "principle" has a factual content since it refers to bodies or particles. But it has no empirical content because it holds only in the absence of any action upon the body of interest, in particular if no empirical tests are performed on it: in short, it does not refer to any empirical situation.

Scientific theories are factual: they concern (authenticated or putatively) real things with properties that are mostly beyond the reach of perception though not of intellection. Therefore they rarely contain empirical concepts. Yet, if they are scientific, theories can be anchored to experience by adding to them links between some theoretical concepts and some concepts having an empirical counterpart. For example, electrodynamics does not deal with pointer readings but gets its support from such operations by virtue of certain formulas that link the field intensity to the torque in a meter. Likewise to test his theories the psychologist looks for behavioral and physiological manifestations of mental phenomena. The physician uses a whole battery of symptoms or diagnostic signs (among them the "vital signs") to evaluate the state of health of his patients. And the economist employs an array of economic and social indicators, such as the steel production and the rate of unemployment, to estimate the state of the economy. All such observables are conjectured, rightly or wrongly, to be lawfully linked to unobservables, so that a measurement of the former allows one to infer (sometimes calculate) the value of the latter. See Figure 11.3.

These observable-unobservable links used to be called *operational definitions* (Bridgman, 1927). But, since they are hypotheses to be tested, not conventions, they had better be called *indicator hypotheses*. An indicator hypothesis is a hypothesis relating an unobservable property of a thing to an observable property of the same thing or of a second thing lawfully linked to the first. The hypothesis may be precise or imprecise, according as it is formulated mathematically or not, but in any case it must

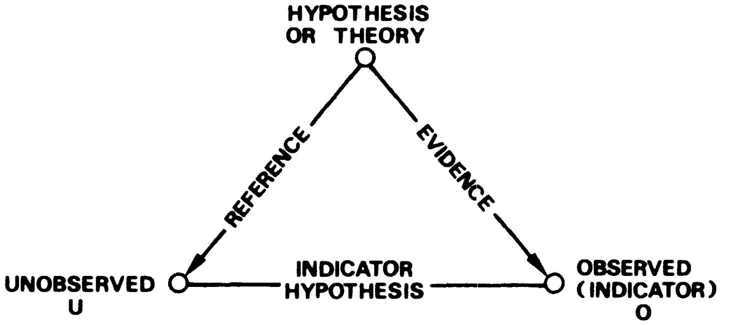


Fig. 11.3. The unobserved-observed (or indicator) relation is expressed by an indicator hypothesis allowing us to infer unobservable things, properties, or events from observation. (Adapted from Bunge (1967a).)

be fully testable and may therefore be improved or replaced with a better one. Far from replacing theoretical concepts, indicator hypotheses are added to factual theories in order to render them testable. Such enrichment of theories in preparation for their empirical tests is sometimes called *operationalization*. A hypothesis or theory that cannot be operationalized, or cannot be logically linked to any operationalizable constructs, is sheer speculation and therefore does not qualify as scientific or technological.

Many, perhaps most of the indicators used in ordinary life and in some professions are imprecise and empirical and therefore unreliable. The cook takes a potato at random from the pot and pricks it with a fork to find out whether it is boiled; but the potato could be rotten, or it could be the only one from the new crop. The naive person takes certain gestures of another as tokens of sincere friendship, while actually they may have been calculated to produce this impression. He is using incorrectly the true generalization "For every x and y, if x is friendly towards y, then x gives y tokens of friendship" to infer friendship from a smile, a kind sentence, or a gift. He is of course indulging in a logical fallacy. A more experienced person would wait for the outcome of further tests; and in future there may be precise philometers.

Typical forms of ambiguous indicator hypotheses are *If U then O*, and *If O then U_1 or U_2*, where O stands for an observable property and the U's for unobservables. In the first case the presence of an indicator (observable) is only a necessary but not a sufficient condition for the corresponding unobservable. In the second case the indicator points ambiguously to two (or more) unobservables. In both cases the indicator–unobservable relation is one–many rather than one–one, and therefore unreliable. Thus the long backlog of a scientific journal may indicate either its prestige, or the scarcity of outlets in its field, or a sudden swell of interest in it, or the poverty of the supporters of the journal, or a deficiency in its administration. Only an independent inquiry may discover the cause(s) of the backlog. Another example: if a political candidate is beaten at the polls, this may indicate that he was not well known or too well known ;that he spent too little money or too much; that he was regarded as too dumb or too smart; that he represented the interests of a minority or that most people failed to understand that he did represent them—and so on.

There are two ways of dealing with ambiguous indicators. One is to multiply them, i.e. to employ a whole battery of mutually consistent indicator hypotheses, hoping to reduce the uncertainty. For example, a physician observes symptom O_1, which may indicate causes U_1 or U_2. But he also observes O_2, which never accompanies U_2, so he infers that U_1 is the cause. Even so he may order a few laboratory tests, not only to make sure that U_1 is the cause but also to estimate its strength. Whereas some of these tests are direct, others are indirect, i.e. they yield further indicators. Thus an X-ray plate is direct because it images the suspect bone or organ. On the other hand most of the results of blood tests are indirect.

A second way of reducing uncertainty about unobservables is to conjecture and test one–one indicator hypotheses, i.e. to substitute 1:1 functions for relations. If we suspect or have established that the observable O is a certain 1:1 function f (e.g. a linear function) of the unobservable U, then by inverting f we may compute U from the measured O values, for $U = f^{-1}(O)$. In this case we say that *O measures U*. (Caution: indicators measure unobservables, but only an effective measurement operation can pin numbers on either.) Needless to say, it is desirable, though not always possible, to have different measures of every important unobservable, so that one may check the other.

Here are a few simple examples of functional hence unambiguous indicator hypotheses. (*a*) The angle of deviation of the pointer in an ammeter connected in series measures the intensity of the electric current in a circuit. (*b*) The intensity of radioactivity can be measured by the discharge of charged bodies, for this discharge is caused by the ionization of the gas surrounding the radioactive sample. (*c*) The quantity of ACTH (adrenocorticotropic hormone) released is an indicator of the activity of the adrenocorticotropic system, which in turn is a measure of stress. (*d*). The time at which two related species diverged is proportional to the fraction of number of kinds of protein that they fail to share: this hypothesis is the basis of the "molecular clock" method for measuring inter-species distances. (*e*) The average length of a bus queue is proportional to the time elapsed since the last bus stopped by.

Whereas some indicator hypotheses have sound theoretical bases and have been abundantly confirmed, others are tentative conjectures with insufficient empirical support. The hypothesis (*a*) underlying all ammeter readings is of the first kind, whereas the hypothesis (*b*), on which the "molecular clock" method is based, is of the second kind. Should rapid speciation events, taking between 5,000 and 50,000 years (like the one reported on by Williamson (1981)), prove frequent, then hypothesis (*b*) is likely to hold only, if at all, for very long periods, of the order of one million years. Another indicator hypothesis that has recently been challenged is the one according to which genotypes are mapped onto phenotypes in a one to one fashion, so that, given a genotype, the corresponding phenotype can be inferred, and conversely. This simple and widely accepted hypothesis is false. First, there are the cases of evolutionary convergence, showing remarkable morphological similarities (e.g. between cetaceans and fish) hiding profound organismic and therefore genetic differences. Second, there are species (or superspecies) within which an extensive karyotic diversity is found underneath little morphological differentiation. (A striking case is that of the spiny rats from Venezuela, the number of chromosomes of which varies between 24 and 62: Benado *et al.*, 1979.) In short, the genotype-phenotype relation is many to one, which renders phenotype a highly unreliable genotype indicator.

Reliable indicators help operationalize a theory, i.e. prepare it for empirical tests, in the following manner. Call $\mathcal{T}$ the general theory to be tested and S the set of subsidiary assumptions specifying particular features of the referent (e.g. the composition, environment, and structure of a system). From $\mathcal{T}$ and S we build the (bound) model $\mathcal{T}'$ of the referent, which is to be subjected to tests. Enter the set I of indicators built with the help of the antecedent knowledge and $\mathcal{T}$ itself. Enter also the set E of available empirical data concerning the referents of the theory and relevant to the latter. Such data are somewhat remote from the theory: for example they may consist of behavioral information, whereas the theory is about mental states. This theory-experience gap is bridged by the indicator hypotheses I: they allow us to "read" behavioral changes in terms of mental processes, benefits from costs and sale prices, or what have you. We may call this a *translation* of the available data into the language of the theory. These translated data E', that follow logically from E and I, are fed into the model $\mathcal{T}'$ to yield the translated model $\mathcal{T}''$. Finally, this result is

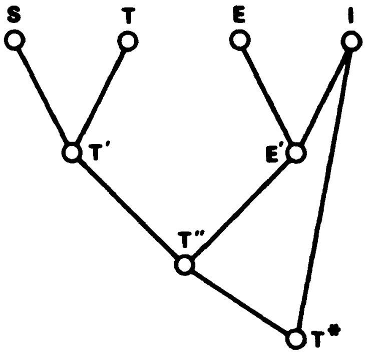


Fig. 11.4. The operationalization of a theory T. T and the specific assumptions S yield the model T' of the referent. The data E are translated into the language of the theory with the help of the indicator hypotheses I: the result is E' which, together with the model T', yield the testable consequences T''. Finally, these are translated back into the language of experimental science by means of the same indicator hypotheses I, to yield T^*, which is ready to be confronted to fresh data.

translated back into the language of experience by means of I. That is, $\mathcal{T}''$ is joined to I to entail $\mathcal{T}^*$, the operationalization of the theory, or rather of the (bound) model $\mathcal{T}'$. Hence not $\mathcal{T}$ itself but some consequences of $\mathcal{T}$ together with the subsidiary assumptions S, the data E, and the indicator hypotheses I, face whatever fresh experience may be relevant to $\mathcal{T}$. See Figure 11.4.

Our view of theory testing is at variance with the received view among philosophers of science. According to the latter a theory is a hierarchy of hypotheses, the highest placed of which would contain theoretical concepts without a directly observable counterpart. The lowest level hypotheses, instead, would be mere generalizations from observable facts and would therefore contain only observational concepts. And there would also be mid-level hypotheses containing concepts of both kinds. Moreover the relation between hypotheses of different levels would be strictly logical: the higher level hypotheses would imply the lower level ones. Therefore, theories would be tested by checking their lower level components, i.e. the logical consequences of their axioms. Such consequences would be in finite numbers because they are allegedly the inductive generalizations that prompted the construction of the theory.

This simple theory of theories is false, for it assumes that theories contain empirical or observational concepts in addition to theoretical ones, whereas in fact empirical concepts occur only in the experimental protocols, and even there combined with theoretical concepts. Also, it is not true that deduction can perform the feat of taking us from the theoretical or abstract to the empirical or concrete. What is true is that the axioms of a theory are seldom independently testable because no axiom contains all the predicates necessary to identify a thing, state, or event. The postulates of a factual theory must be combined with one another, as well as enriched with additional hypotheses, data and indicator hypotheses, to yield predictions concerning specific things in specific states. In short, we must inject some empirical content into a theory if we wish the theory to face experience: by itself it has none.

To conclude, indicator hypotheses are indispensable to test hypotheses and theories. But many indicators used in ordinary life cannot be employed in science or technology because they are ambiguous, imprecise, or both. Not all laughter indicates joy, and not every fluctuation in the Dow-Jones index represents a real change in the economy. The search for reliable indicators is a task both theoretical and empirical. The first because only a theory can tell us whether a given sign does in fact point to a certain unobservable rather than another, by disclosing the objective link between observable and unobservable. And only empirical checks can validate an operationalization procedure, for it involves hypotheses.

3. DATA

3.1. *Measurement*

We need objective and precise information about the world, i.e. accurate data, for a number of purposes: to conduct the business of daily life, plan our activities, check our hypotheses and theories, and to explain and predict with the help of the latter. There are two main ways of learning data: producing them oneself or borrowing them from others. And there are three modes of generating data: by observation, measurement, and experiment. Observation, whether direct of with the help of instruments and theories, is deliberate and controlled perception, and it is the basic mode of data generation. We studied it in Ch. 4, Section 2. We shall now examine measurement, which may be characterized as quantitative observation, or the observation of quantitative properties. Experiment, or the observation (and possibly measurement) of changes under our partial control, will be tackled in the next subsection.

All organisms, even unicellular ones, have sensors capable of detecting and even gauging certain changes, in particular gradients and deviations from the optimal values of certain parameters, such as temperature, salinity, and acidity. But only man has invented and constructed, over the past few millennia, a variety of measuring instruments of varying precision. The differences between these and the natural "measuring" devices are as follows: (*a*) every measurement proper presupposes the conceptual operation of quantitation, or assignment of numbers to the degrees of a property—e.g. the conversion of separation into distance; (*b*) some measurements presuppose also the construction of indicator hypotheses—e.g. that the angular velocity of the anemometer measures the speed of wind; (*c*) measuring instruments are equipped with pointers, digital dials, or some other indicators, as well as with scales, that allow one to "read" them; (*d*) the measuring instruments are detachable, they can be exchanged for others, repaired, and improved on. We studied quantitation in Ch. 5, Section 3.2, and indicators in the last section. We shall not study measuring instruments, the subject of instrumentation (see, e.g. Stein, 1965); instead we shall discuss some of the methodological and philosophical issues raised by measurement.

The very first such problem is: *What is measurable*? The usual answer is that facts are measurable for, according to a well-known formula, measuring is "pinning numbers on facts" (S. S. Stevens). What kinds of facts? Surely not things but their properties and changes thereof. And which ones? Obviously only the ones that can be manifested objectively—e.g. generous actions but not (so far) generous feelings or intentions; and prices but not values (in use). Clearly, but what is a measurable property? The short answer is this: A property is *measurable in principle* iff it is (*a*) quantitative (i.e. the corresponding concept has been quantitated) and (*b*) either manifest (observable) or lawfully related (through an indicator hypothesis) to a manifest property. Whether a property is actually measurable at a given time depends on the state of the art, opportunity, and resources.

The two measurability conditions above, (*a*) and (*b*), can be summed up in the phrase 'theory dependence'. This dependence is denied by empiricism, according to which measurement data belong to the theory-free "empirical basis" of knowledge. Experimentalists know better: they know that the very design and operation of measurement instruments depends on theories. For example Rutherford, perhaps the greatest experimentalist in our century, organized as follows the first part of his classic treatise *Radio-Activity* (1904). He devoted the first chapter to a qualitative description of radioactive substances, the second to the *theory* of ionization of gases, and the third to measurement methods (some of which use that theory, and others different theories). This was necessary to explain why the discharge of charged bodies is a reliable indicator of the presence of a radioactive substance (the radiation of which ionizes the surrounding air). In the absence of some well-confirmed theory one may "pin numbers on facts" without understanding what one is measuring. IQ measurement is a case in point: it is not based on theory and it can be put to evil uses. (See Gould, 1981.)

If a theory states that objects of kind K have property P, then the experimenter may decide to devise a procedure for measuring P. (For example, unlike action at a distance theories, the electromagnetic field theories prompted experimentalists to produce and detect electromagnetic waves, and to measure their properties, such as wavelength and frequency.) If the experimentalist fails to detect the theoretically predicted property, then either the theory is at fault or his experimental design is inadequate. Should the theory be successful in a number of other cases, the latter inference is in order, and alternative experimental designs ought to be tried. If on the other hand a successful theory does not state that the objects of kind K have property P, then if the theory is true, P must be regarded as unmeasurable, perhaps because nonexistent.

For example, according to quantum mechanics the three components of the spin of a particle cannot all be measured at the same time. (This is how the noncommutativity among the spin components is usually interpreted. In my own interpretation unmeasurability derives from nonexistence: if one of the spin components is "well defined", i.e. has a sharp value, then the other two are not.) On the other hand, according to a classical or semiclassical theory, where the spin and other dynamical variables are "hidden" (i.e. scatter-free) variables, the particle would possess all three components at the same time, so they would be simultaneously measurable in principle. In short, measurability is critically dependent upon theory.

The central philosophical and methodological question about measurement is of course: What is it? The short answer is poetic rather than technical: Measurement is a synthesis of fact, reason, and action. Let us spell this out. (See Pfanzagl (1959) and Bunge (1967b) for different versions.) Consider a thing possessing a property P that comes in degrees, and call $\mathscr{P} = \langle P, \lesssim \rangle$ the relational system formed by the set of such degrees together with its intrinsic factual order $\lesssim$ (e.g. weaker than or as strong as). Conceptualize $\mathscr{P}$ by constructing a magnitude the values of which are, in the simplest case, real numbers together with units—e.g. 7 cm $\sec^{-1}$—and call $\mathscr{R} = \langle R, \leq \rangle$ the set of such values together with its order $\leq$.

Quantitation (or numerical quantification) is a mapping of $\mathscr{P}$ into $\mathscr{R}$ preserving (representing) the objective order of $\mathscr{P}$.

Look now at the set M of marks on a meter designed to measure the given property via some indicator hypothesis—e.g. in such a manner that *weaker* corresponds to *to the left of.* Call $\mathscr{M} = \langle M, \lesssim \rangle$ the new factual relational system, where now $\lesssim$ is the conventional order of the marks. The measuring device involves a one to one correspondence between $\mathscr{P}$ and $\mathscr{M}$, and between $\mathscr{M}$ and some subset Q of the rational numbers (fractions) together with the same units as before. Some of the latter numbers will be represented on the dial of the meter. Call $\mathscr{Q} = \langle Q, \leq \rangle$ the conceptual relational system formed by such numbers (together with the appropriate units) in their natural order.

Altogether we have two *factual* relational systems, $\mathscr{P}$ and $\mathscr{M}$, and two *conceptual* ones, $\mathscr{R}$ and $\mathscr{Q}$: see Figure 11.5. These four systems are related as shown in the following diagram:

$$\begin{tikzcd}
\langle P, \lesssim \rangle \arrow[r, "Quantitation"] \arrow[d, "Indication"'] \arrow[dr, "Measurement" description] & \langle R, \leq \rangle \arrow[d, "Inclusion"] \\
\langle M, \lesssim \rangle \arrow[r, "Scaling"'] & \langle Q, \leq \rangle
\end{tikzcd}$$

This diagram answers our question, albeit in a somewhat crytic manner, *viz*: Measurement is the empirical operation that maps some of the objective degrees of a property into rational numbers (together with the appropriate units) via indication and scaling.

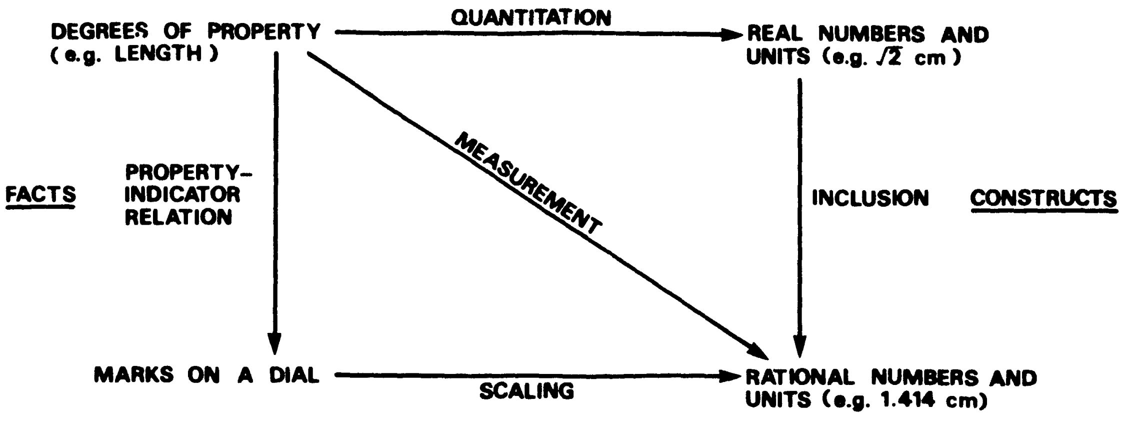

DEGREES OF PROPERTY
(e.g. LENGTH)
QUANTITATION
REAL NUMBERS AND
UNITS (e.g. √2 cm)
FACTS
PROPERTY-
INDICATOR
RELATION
MEASUREMENT
INCLUSION
CONSTRUCTS
MARKS ON A DIAL
SCALING
RATIONAL NUMBERS AND
UNITS (e.g. 1.414 cm)

Fig. 11.5. Measurement as a one to one correspondence between the objective degrees of a property and instrument readings. Modified from Bunge (1967b, p. 221).

What is being measured depends not only on the object of measurement but also on the measurement technique, which includes instrument design. Two or more techniques may or may not measure the same magnitude, and it is only when they do that we must require that the result obtained with their help exhibit a strong correlation. The case of speed measurement should make this point clear. An observer riding a train has at least three different methods of measuring the train speed, each of which yields values of a different function (Lévy-Leblond, 1980). First there is the external procedure: the observer counts the mileposts placed regularly along the track, and reads the times shown by the clocks in the railway stations. He thus obtains a value of the *velocity* or rather its average. Secondly, there is the mixed procedure: the observer still counts milestones but measures time on his own watch. What he measures now is a different property, namely *celerity*. A third procedure is purely internal, namely to employ an accelerometer and plot its values against the proper time of his watch. This allows him to measure the train's *rapidity*. For slow motions the values of all three functions are the same, but for fast ones they are very different. To take a simpler example, each aspect of aphasia is studied by a different method even though they are all grouped under the same heading. Thus the sodium amytal procedure tests speech, the tachistoscopic one reading, and the dichotic measure listening (Herron *et al.*, 1979). More on measurement techniques, in particular instrument design, in Bunge (1967b, Ch. 13, Section 13.5).

The point of the above examples was to stress the platitude that different measurement techniques *may* measure different properties. Operationism has stretched this point claiming that different measurement procedures *always* "define" different concepts (Bridgman, 1927)—even if the theory does not make such distinction. (A classical example, still found in books, is the pseudodistinction between inertial and gravitational mass. For a criticism see Bunge (1967c).) But of course only theories can draw conceptual distinctions; and if a theory states that two allegedly different concepts are one, then different measurement methods can only give (different or equal) values of one and the same property.

Measurement, then, does not define concepts, let alone things. Instead, it helps us find out some of the properties of things and in this way identify them correctly or control them efficiently. For example, there are alternative procedures for identifying acids, the litmus paper technique being the most popular. Each such procedure can be summarized into an operational *criterion* (not definition), such as "Anything that turns litmus paper red is an acid". Unlike a definition, such a criterion tells us not what an acid *is* but how to *recognize* it. Therefore operational criteria cannot replace theory; moreover they can be justified only by theory. Of course, in turn theories must be justified with the help of empirical operations, in particular measurements. So, we have a feedback loop or virtuous circle: "There is no measurement without theory and there is no theory without measurement" (Stein, 1965, p. 5).

What holds for properties holds also, a fortiori, for their lawful relations: an effort should always be made to try and check the latter by alternative methods for measuring the same properties, even at the risk of incurring the disapproval of short-sighted administrators who may regard such checking operations as wasteful. A strict operationist might object that different methods bring about different relations among properties, i.e. different laws. This claim has actually been made with regard to the psychophysical law relating the sensation intensity to the stimulus intensity (Wasserman *et al.*, 1979). What may have happened in this case is that differences in method elicit differences in the internal states of the experimental subject, which states are ignored by the law in question. This is the rule rather than the exception. If method 1 allows one to check the functional relation "$y = f_1(x)$", and method 2 "awakens" a second independent variable z and establishes the relation "$y = f_2(x, z)$", then the experimenter is likely to rejoice for having discovered a more inclusive relation. And he will be even more satisfied if he succeeds in showing that, upon "freezing" z (or restricting z to varying within a narrow interval), f_2 comes close to f_1, which is then shown to be but a first approximation to f_2.

A related philosophical problem is whether all measurement creates its object or at least interferes with it and whether, when it does, such interference can be corrected. A measurement technique is said to be *obtrusive* if it changes in any way the state (and a fortiori the kind) of the object of measurement, and *unobtrusive* otherwise. *Example 1*: A current or voltage measurement is slightly obtrusive, for it involves dissipating a fraction of the energy in the meter. But this effect can be corrected for with help of a second measurement, so the net effect is negligible. *Example 2*: Measuring the age of a fossil using the carbon 14 dating technique is highly obtrusive, for it requires burning the object. But alternative dating techniques have been invented which are far less obtrusive. *Example 3*: All interviews and questionnaires used by social scientists and psychologists are somewhat obtrusive, for people change their behavior when they know that they are being observed. Think of questions such as "How often do you batter your children?", "Are you the kind of people who help their neighbor in need?", and "Do you sympathize with communism?"

There is no question but that, the deeper and more accurate a measurement technique, the more obtrusive it is likely to be. However, the history of instrumentation shows that, if we try hard, we are likely to invent means to minimize the effect of the observer's interference or at least to calculate its effect. For example, by lowering the temperature to near the absolute zero, thermal noise is eliminated, and magnetic fields are kept out of metals; and, by using semitransparent mirrors and hidden cameras, the experimental subjects cease to act in a self-conscious or in a hypocritical manner.

Still, there is one important case where it would seem that obtrusiveness is here to stay, namely in the measurement of properties of microthings. For example, if the position of an electron is measured with great accuracy, then—according to the Heisenberg inequalities—it is impossible to measure its velocity at the same time. However, these formulas can be interpreted in several ways (Bunge, 1977c). According to the popular interpretation, the measuring device disturbs the electron in an unpredictable way, imparting it an uncontrollable momentum: this is the *causal* intepretation, which ill accords with the stochastic nature of quantum mechanics. The Copenhagen interpretation avoids causality but introduces a subjectivity alien to the scientific approach: it states that the electron *has* no properties as long as these are not measured, and acquires them the way the experimenter wants.

The realist interpretation is that, in general, the electron *has* neither a sharp position nor a sharp momentum but only distributions of both, as well as the possibility of *acquiring* either a sharp position or a sharp momentum, but not both at the same time, under suitable environmental conditions. The electron position *becomes* measurable only if the electron is forced, by appropriate natural or artificial means, to adopt a state characterized by such property—which it will do only at the price of losing whatever sharp velocity value it may have had before. And if we decide instead to measure the electron velocity we must start by preparing the electron in a state characterized by a sharp velocity value rather than by a whole range of velocities. We cannot measure a property that the electron does not possess to begin with. There is nothing subjective about all this, for the conditions under which an electron gets its position (or velocity) sharpened may be natural or artificial. And, contrary to popular wisdom, the Heisenberg inequalities do not impose any limitations upon our knowledge: they only show that microthings are not particles. (But we should have known this ever since classical mechanics was shown inadequate to describe, explain and predict their behavior.) In short, we certainly alter the state of a microthing when measuring some of its properties, but we do not thereby create the thing at our whim and in violation of the laws of nature. When an experimenter undertakes to measure a property on an object he assumes the objective existence of some objects possessing certain properties and laws.

Still, it might be thought that, since natural things do not come with numbers indicating the values of their properties, the outcomes of measurement are wholly man-made. After all, by shifting the zero of a scale or adopting a different unit we may assign a thing any size or age. Likewise empty space is devoid of milestones, so why not think that it is metrically amorphous, hence that our assigning it metrical properties is purely conventional? The answers to these conventionalist innuendos are as follows. First, the pinning down of numbers on properties is basically no different from our constructing theories about things. Surely Maxwell's equations are not written on the electromagnetic waves that illumine this book, but those equations do fit them quite well. Second, we often handle, and in principle could always deal with, ratios—e.g. relative concentrations, densities, and the like—rather than with absolute quantities, and such ratios are unit-free, hence independent of any unit conventions. Thirdly, although the choice of scale (in particular zero and unit) is conventional, such choice, when adhered to consistently, renders intersubjective comparison and communication possible. Language too is largely conventional, yet such conventions are necessary for communication.

Units occur explicitly in the very formation of dimensional magnitudes such as permeability and productivity. (Cf. Ch. 9, Section 3.1.) All scientists are familiar with the rules for the formation of such concepts, and will reject as ill formed an expression such as 'The mass of body *b* equals number *c*', admitting instead expressions of the form 'The mass of body *b* equals *c* grams', or 'The mass of *b* in grams equals *c*'. (The classical mass concept is formalized as a function from pairs $\langle$body, mass unit$\rangle$ to positive numbers.) Units are defined in conceptual terms, preferably in terms of law statements. For example the newton, a force unit, is defined via Newton's second law of motion. (One newton is the force that causes an acceleration of one meter per second per second to a body possessing a mass of one kilogram.)

In addition to units we need standards for some of them, i.e. their materializations. Such standards are carefully described physical things or processes that can be easily replicated or reproduced as well as improved on. Thus the length standard used to be a meter long platinum bar under certain pressure and temperature conditions. As from 1983 the meter standard is the distance light travels in vacuum in a given fraction of a second. (The speed of light is taken as a base line equal to 299,792,458 m/sec.) Contrary to careless textbook affirmations, standards do not *define* units but *materialize* them. And, unlike units, which are arbitrary, standards are chosen for their stability and reproducibility. Entire governmental laboratories, such as the US Bureau of Standards, are devoted to keeping and replicating standards as well as to investigating ways to improve on them.

Needless to say, improvements in standards and, in general, in instrumentation, require imagination and motivation allied to familiarity with basic theory, electronics, and occasionally also computer programming. Improvements in instrumentation over the past few decades have often advanced at the rate of one order of magnitude (i.e. a factor of 10) every 5 years or so. Fewer and fewer institutions can afford to keep this pace. Hence, unless something drastic be done, experimental research will soon become the privilege of a few giant industrial laboratories—which are not interested in basic research anyway. There is then as much to bemoan as to celebrate in the swift progress of instrumentation.

In addition to favoring an excessive concentration of human and material resources in a few rich laboratories, quick progress in instrumentation may have the following negative side effects: (*a*) an accumulation of detail may make us forget the whole; (*b*) observation, measurement and computation, which are only means, may be turned into ends, and correspondingly scientists into technicians; (*c*) alternative approaches, requiring more modest laboratory equipment, tend to be ignored as the value of research projects tends to be gauged by their cost; (*d*) when too much reliance is placed on instruments, basic questions are hardly asked, and ambitious theories hardly sought. Austerity has its virtues: as Lord Rutherford used to say, "We've got no money, so we've got to think."

Suppose now that all the preliminary conceptual operations have been performed: we have constructed or adopted the magnitude (quantitative attribute) to be measured, and if necessary also its hypothetical (though reasonably well confirmed) relation to an accessible property (indicator).

Moreover we have designed our equipment with the help of a bunch of law statements, and have checked and calibrated it, so that it is ready to measure, to within an acceptable error, what we want to measure. Moreover, suppose we have obtained a suitable preparation—crystal, insect, or what have you—known or suspected to possess the property of interest. (This is no mean task, for we want a specimen or a sample representative of the original population, relatively easy to study, and we wish to minimize any destructive interference with it.) Only now are we ready to perform the measurements and record them, or perhaps to entrust them to a computer-assisted device. Quite frequently, whereas this last stage may be completed in a matter of days, the previous stage may consume months or even years of work sometimes involving dozens of coworkers and technicians.

We now have a set of raw data. They are not sacrosanct: they must be screened, for some of them may be far too weak signals and others plain errors. The remaining data are corrected in some way or other—e.g. reduced to standard atmospheric pressure and temperature. Even this data reduction may be insufficient: usually we want to extract from them some characteristics representing them as a whole, such as the average and the scatter around it (i.e. the standard deviation). The latter may be taken as a good measure of the measurement error. Such error is not necessarily a mistake on our part. Sometimes it exhibits the spontaneous random (e.g. thermal) changes of both object and apparatus. (It can be considerably decreased by working at very low temperatures.) At other times the statistical scatter exhibits the variety of the individuals composing the sample under investigation. (In sum, σ is an ambiguous indicator.) In any case, the correct summary of a run of measurement reads thus: The average value plus or minus one standard deviation.

The calculated average is usually called the "real" or "true" value of the property of interest although actually every one of the values that goes into the calculation is real. Ideally we should keep them all, or at least we should not throw away all of them unless we can make sure that the sample under study is rather homogeneous in the respect of interest. See Figure 11.6 for a case of hasty throwing away of valuable information. In any event, the results of measurements are "cooked" before being compared with those of other runs or with theoretical values. Mathematical statistics and the theory of errors of observation study such data "cooking" procedures and pose a number of genuine and interesting methodological problems that are still awaiting philosophical analysis. Suffice it to mention the ground, or

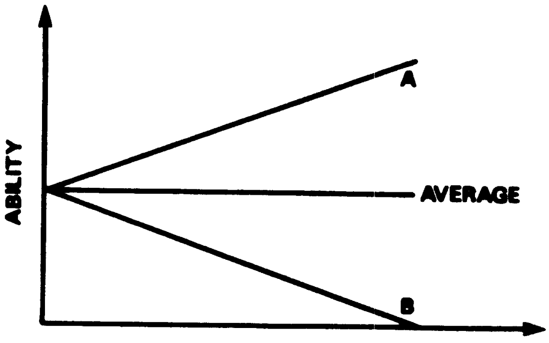


Fig. 11.6. A group of experimental animals is tested for some ability and its variation in the course of time. Those of kind *A* increased their ability while those of kind *B* decreased it. The average showed no change, so if it had been taken as a faithful representative of the group it would be uniformative. Adapted from R. M. Yerkes (1907) *The Dancing Mouse*, p. 273. *New York: Macmillan.*

lack of it, for adopting averages as the real or true values, and for discarding outlying values (i.e. measurement results departing from the average by more than three standard deviations).

We close this section with some remarks on two supposedly extremely general theories of measurement. One of them is what mathematical psychologists call *measurement theory* (e.g. Stevens, 1951; Suppes and Zinnes, 1963), and which others would call 'formation and analysis of scientific (in particular quantitative) concepts'. In fact this theory is unconcerned with the empirical operations everyone else calls 'measurements'—such as the actual determination of durations and electric charges—and deals instead with what it calls 'measurement scales', and the adult sciences call 'quantities' or 'magnitudes'. The theory contains basic mistakes that render it useless. One of them is that it ignores dimensional analysis and even the very notion of a dimension. Another is that it does not contain a proper account of units. (For algebraic theories of dimensions and units see Bunge (1971).) A third is that it ignores intensive quantities, such as density and *per capita* income, which are often primitive in scientific theory. A fourth is that is postulates that all extensive magnitudes (such as length) are additive, whereas in fact some of them, such as mass and entropy, are subadditive. (I.e. the mass of a system is somewhat smaller than the sum of the masses of its components.) The moral is plain: It is not advisable to try and build general theories of scientific concepts with disregard for the factual theories that house such concepts. (For details see Bunge (1973c).)

Another purportedly general theory of measurement is the *quantum* *theory of measurement* or, rather, the family of theories dealing with measurement in general on the basis of quantum mechanics. The aim of such theories is to analyze micro–macro interactions, in particular those between a microthing (e.g. an electron) and a measuring apparatus. Whereas some such theories are strictly physical or objective, others make room for the observer's mind (without however containing a single variable representing mental functions). The main trouble with all these theories is that they are nothing but academic exercises consisting in applying quantum mechanics to extremely idealized situations: they ignore that every actual measurement measures one or more *particular* properties on *specific* things with the help of *special* devices.

There is no such thing as the all-purpose measuring instrument imagined in those theories. Since every type of measuring instrument is quite specific, it requires its own theory, which is usually a classical one and occasionally a mixture of classical and quantum theories. (Think of the enormous difference between optical spectrographs and mass spectrographs, or between thermocouples and scintillation counters.) A general theory of measurement may explain some features of measurement, such as its irreversibility and amplification. But, because of its generality, it can issue no precise predictions and, not being predictive, it is empirically untestable. For this reason none of the many general quantum theories of measurement can ever be checked experimentally. So, the choice among them is a matter of indifference: one may choose either of them or be more parsimonious and ignore them all. On the other hand it is legitimate to try and unravel the main features of measurement and even to build a general theory of experiment involving no particular laws and therefore having no predictive power at all: See Appendix 5.

3.2. *Experiment*

Like measurement and unlike mere perception, experiment is active experience and one blended with reasoning. And, like measurement, experiment can find new facts or test hypotheses. However, experiment can find facts that measurement cannot unless it is part of experiment—namely facts that occur because of the deliberate alteration of the object of study. Indeed a fact finding experiment helps solve problems of the form "What happens to X if Y is done to X?"—e.g. what happens to a frog upon the unilateral destruction of its labyrinth, or to a community deprived of its leadership? Experiments of this kind are more difficult to perform than experiments aiming at testing hypotheses of the form "If Y is done to X, then Z results", for in this case the hypothesis itself directs us to look for effects of kind Z when controlling variable Y, whereas in the former case we must look for every possible difference that Y may make.

Unlike measurement of the unobtrusive kind, every experiment involves controlled changes in the object of study. Indeed an experiment may be defined as a controlled action performed by an animal on some other object, or on a part of itself, with the aim of solving some problem—cognitive or practical—concerning the object of experimentation, and such that the animal is able to receive, record, and analyze the reaction of the object to that action. What is typical of all experiment, in contrast to observation, is that the experimenter controls the object and its environment. In some cases this control is limited, i.e. restricted to the properties thought to be of interest. But in others the control is nearly total, for it includes the production of the thing or event of interest—as is the case with experiments with particle accelerators, with the artificial assembly of macromolecules such as genes, and with experimental social groups.

More precisely, an *experimental device* (or *set up*) is a concrete (material) system with at least three components: the object w of study (which can be replaced from one experiment to the next), an object supplying a stimulus (input) of some kind to w, and an object measuring the response (output) of w to that stimulus. An *experiment* on w consists in (*a*) subjecting w to a stimulus and observing its response; (*b*) observing the output of the same object w or one of the same kind when not subjected to the given stimulus, and (*c*) comparing the responses and finding out whether the difference between them is significant, i.e. due to the stimulus rather than attributable to chance or to idiosyncrasies of the object of study.

Needless to say, the choice of object and stimulus, as well as the design of the set up, in particular measuring device, are guided by hypotheses; and the final comparison (allowing one to draw "conclusions" about the effect of the stimulus) is controlled by statistics. The net outcome of an experiment is summarized in a set of statements, tables, or figures called the experimental *finding*. Sometimes nothing is found except that the experiment yields no evidence for a given hypothesis—e.g. that there are freely roaming spirits. Papers reporting on such "negative" results are regularly published in the scientific literature—seldom in the pseudoscientific (e.g. parapsychological or psychoanalytic) journals.

The advantage of experiment over observation and even measurement is best appreciated by studying some cases. *Example 1*: To find out whether information can be transferred from one animal to another by surgical transplant, Lindauer and his coworkers (Martin *et al.*, 1978) proceeded as follows. A collection of honeybees was divided into two groups, the experimental or donor bees and the recipient ones. The donor animals were fed between 10 a.m. and 12 a.m. for six days in a row, whereas the recipient bees were fed either continuously or a variable times, so that they did not learn any feeding pattern. The mushroom bodies (one on each side of the brain) of the donor bees were then transplanted into the ignorant bees. It was found that after 3 or 4 days the recipient bees behaved like the "learned" ones. The conclusion was that some information gets "engrammed" into the brain, and can be transferred by surgical transplant. *Example 2*: To test the hypothesis that speech in right-handed persons is a function of the left cerebral hemisphere alone, Kinsbourne (1971) performed the following experiment. Three right-handed men who had suffered left hemisphere strokes causing aphasia were given injections of amytal, first in the left carotid, then in the right one. The former had no effect, whereas the injections in the right carotid arrested completely what little speech the patients had been left with. The conclusion was that the right ("minor") hemisphere does participate in speech production. *Example3*: To check the hypothesis of the genetic origin of intelligence (held by A. R. Jensen, R. J. Herrenstein, H. J. Eysenck, and others), Schiff *et al.* (1978) compared the school performance of middle-class children with that of working-class children adopted by middle-class families. (Theirs was an *ex post facto* experiment: the foster children were unwitting experimental subjects.) They found that the children perform in school according to their homes, not their genetic origin: i.e. what counts most for learning are the cultural, not the biological parents.

We are so used to the idea that experiment is a necessary component of science and technology that we tend to forget how recent it is, and take it naively for granted that every widely advertised theory, artifact, or procedure, has passed suitable experimental checks. Experiment, as distinct from intuitive trial, was introduced only in the 17th century. Even in the next century physicists made few experiments; most of the empirical support for the new mechanics came from astronomy. Until the end of the 19th century mechanics was regarded as a branch of mathematics, not of physics. Biological experimentation was not conducted systematically until about 1850. Although medicine is supposed to be experimental, the first experiment to determine the efficiency of surgery to deal with breast cancer was not performed until 1981. The first laboratory of experimental psychology was inaugurated in 1879. Even though psychiatry is supposed to be a branch of medicine, the proposal in 1980 by a committee of the U.S. Congress to investigate experimentally the efficacy of psychotherapy raised an outcry among clinical psychologists and psychoanalysts. The very idea of a social experiment is foreign to most people. Very few have heard about the New Jersey negative income tax experiment (1968–1972). And many scientists still have hazy ideas about key features of experimentation. Thus some (e.g. Abraham and Marsden, 1978) regard computer simulation—a valuable auxiliary to actual experiment but no substitute for it—as experimentation. Others believe that any sample, regardless of size and mode of obtainment, will do; and still others commit the gambler's fallacy, believing, e.g. that the sequence HTHTHT of coin tossings is more probable than HHHHHH (Tverksy and Kahneman, 1971). All of which goes to show that insufficient attention is being paid to the methodology of experiment and, in particular, to its conceptual basis.

The design and interpretation of every experiment presuppose a number of hypotheses, most of which are rarely if ever noted, partly because few philosophers have ever analyzed any experiments. The presuppositions of experiment can be grouped into *generic* (shared by all experiments) and *specific* (characteristic of every type of experiment). See Figure 11.7. The generic presuppositions are in turn of two kinds: philosophical and

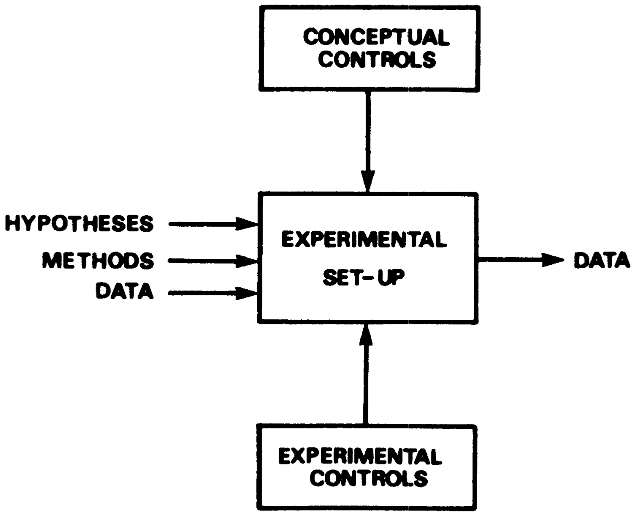


Fig. 11.7. A well-designed experiment has conceptual controls (in particular statistical ones) as well as experimental ones (e.g. of temperature and inflow of air). And in the design and interpretation of the experiment not only data and methods but also hypotheses (philosophical, statistical and scientific) take part.

statistical; and the specific ones consist in particular hypotheses or theories about the nature and behavior of the objects of experiment or the experimental means. Let us examine them quickly.

The main *philosophical* presuppositions of all experiment are these:

(*a*) *Reality*: The members of the experimental and control group, as well as the measuring instruments, exist really although some of the hypothesized objects may be imaginary. (If all the things involved in an experiment were figments of our imagination, then imaginary experiments would suffice.)

(*b*) *Lawfulness*: All the objects involved in the experiment behave lawfully. (There would be no point in performing experiments if nature or society were to give significantly different "answers" every time we pose them the same "question".)

(*c*) *Causality*: All the things involved in the experiment satisfy some form of the causal principle, however weak, e.g. "Every event is the effect (with some probability) of some other event". (Otherwise no deliberate production of an effect and no effective control of variables would be possible.)

(*d*) *Randomness*: All the variables involved in the experiment are subject to some random fluctuation, both intrinsic and due to external perturbations. (Otherwise we would not be able to explain the statistical scatter of results.)

(*e*) *Insulation*: Objects other than the object of the experiment, the experimenter, and his experimental means, can be neutralized or at least monitored for the duration of the experiment. (Otherwise no significant changes could be attributed exclusively to changes in the control variables.)

(*f*) *Disturbances or artifacts*: It is always possible to correct to some extent, either empirically or theoretically, for the "artifacts", disturbances or contaminations caused by the experimental procedures. These are not the deliberate alterations of the object but the unwanted distortions of it or of its image—e.g. the colors and diffuse rings produced by optical lenses, the solvents that change the reactivity of reactants, and the stains that impregnate living tissues with metals. (If such partial corrections were impossible we could not legitimately claim that the thing for us—as it appears to us—resembles closely the thing in itself—such as it is when not subjected to experiment.)

(*g*) *No psi*: It is always possible to design experiments in a manner such that the experimenter's mental processes exert no direct influence on the outcome of the experiment. I.e. the experimenter can be shielded or uncoupled from the experimental set up, so that his bodily and in particular brain processes do not alter the experimental results. Servomechanisms are particularly useful to this end. (Otherwise the outcome of the experiment could be produced at the experimenter's whim, and the experimenter would be testing nothing but his own mental, e.g. psychokinetic, abilities.)

(*h*) *Explicability*: It is always possible to justify (explain), at least in outline, how the experimental set up works, i.e. what it does. I.e. it is possible to form a conceptual model of the experimental device using well confirmed hypotheses and data. (Otherwise we would be unable to draw any conclusions.)

In addition to the foregoing philosophical principles, the following *statistical* principles and rules are supposed to be observed in every experiment, particularly in the biological and social sciences:

(*a*) The objects of experiment should be drawn from as *homogeneous* a population as possible. To this end the original population, if extremely heterogeneous, should be partitioned into equivalence classes. (Otherwise no general patterns may be recognized and correspondingly no generalizations may be conjectured.)

(*b*) When the total population is practically inaccessible (e.g. for being too large), a sample is to be drawn from it in such a way that it be *representative* of the former. Mathematical statistics has come up with techniques for maximizing representativity.

(*c*) The sample to be experimented with should be divided into two roughly equal parts: the *experimental* group (the members of which will be subjected to the given stimuli) and the *control*. The assignment to each group should be done in a random manner, in order to minimize bias in the choice of experimental subjects as well as placebo effects and last, but not least, in order to be able to apply probability ideas. (No randomness, no probability.)

(*d*) Experiments with human subjects should be *double blind* in order to avoid the placebo effect: i.e. neither the subject nor the experimenter should know who gets what. This information should be reserved to technicians or computers and utilized only once all the results of the experiment are in.

(*e*) The experimental design should be *checked* for possible systematic errors (or errors in design), and the accidental (random) errors should be calculated and noted.

(*f*) The *significance* in the differences of behavior between the experimental and the control group should be evaluated with the help of some significance test (e.g. chi-square).

(*g*) The experiment should be evaluated, and if possible repeated, by *independent* groups—not to forestall deception but to correct possible systematic errors.

This incomplete list should suffice to show that, contrary to widespread opinion, statistics should not enter only at the very end, to "process the data", but should be involved throughout the experiment, from design to evaluation. The regulative function of statistics is particularly apparent when the data arrive sequentially rather than in whole batches. In the first case the experimenter can stop gathering or producing data but must do so according to definite rules, for otherwise he may commit the fallacy of optimal stopping, or terminating the experiment as soon as his favorite hypothesis has been confirmed—a habitual tactics of parapsychologists. Such "stopping rules", or prescriptions for deciding when to stop gathering data, are fashioned by sequential statistical analysis.

Some of the above-mentioned statistical safeguards can be ignored in physics and chemistry, which usually work with extremely homogeneous populations (e.g. rather pure chemicals). But the biological and social sciences, which deal with complex systems characterized by an enormous individual variability, cannot ignore them. Yet, although those principles have been taught in standard courses over the past half century, they are still often ignored. *Example 1*: Until very recently studies of mental development and of neurophysiological "correlates" of learning were conducted without adequate controls and sometimes with no controls at all (Denenberg, 1979; Thompson *et al.*, 1972). *Example 2*: Experiments in psychoanalysis are conspicuous by their absence, and in parapsychology notorious for inadequate controls and for a fanciful use of probabilistic and statistical ideas. (Cf. Hansel, 1980; Alcock, 1981.) *Example 3*: The most famous of all experiments in human engineering since the days of Taylor was the Hawthorn experiment (1927–1933) on the effect of human relations on productivity. Its analysis half a century later (Franke and Kaul, 1978) has revealed that it had not been an experiment proper because no control groups had been involved, so that the inferences had been of the *post hoc propter hoc* type. Second, the data were not subject to statistical analysis. Third, between the start and the termination of the "experiment" the Great Depression set in, introducing a major variable which the investigators were not in a position to control, namely the fear of unemployment and the corresponding toughening of management. Fourth, fatigue was reduced by working fewer hours a week.

Finally we come to the *scientific* hypotheses underlying the design, performance, and interpretation of experiments. These hypotheses are of two kinds: whereas some of them refer to the means (in particular the measuring instruments), others concern the objects of experiment and their interaction with the former. We examined the hypotheses of the first type in the preceding section; let us now look into those of the second type. At first sight no such hypotheses are involved in experiments attempting to answer questions of the form "What happens if *X* is done to *Y*?" A closer examination reveals that, far from choosing the stimulus *X* at random, the experimenter is guided by both his information, however scanty, about *X* and *Y* and some hunches, however vague, on the possible effects of *X* on *Y*. Otherwise he would succeed only by sheer luck, for experimenters do not have an infinite time and things are not susceptible to all possible stimuli. (Try to play Beethoven to a fly. Rock music, on the other hand, might have some damaging effect.)

Every well-designed experiment is inspired or misguided by some specific hypothesis or other. For example, an experiment inspired in stimulus–response psychology controls only behavioral variables. Hence it can draw valid conclusions only if the experimenter has "frozen" (kept constant) all the relevant internal variables, such as drives and motivations. (For example, it has been reported that rats prefer complicated roads leading to food sources, to straight ones. Maybe so, but any experiment to test such curiosity or adventurousness is worthless unless it has been ascertained that both the curious and the greedy rats are equally hungry.) On the other hand an experiment in physiological psychology will control physiological as well as behavioral variables. For example, to find out whether neural system *X* is "involved" in (or "mediates") behavior *Y*, block or stimulate the activity of *X*, and watch for behavior of kind *Y*. (*Dual*: Modify behavior from *Y* to *Z*, e.g. by placing constraints on motion or on perception, and watch for changes in the activity of neural system *X*.) Another example: it is usually assumed that it is possible to separate hereditary from environmental factors in the study of intelligence by using the twin method. (Each of the monozygotic twins is placed in a different environment, and their performance is compared.) This technique presupposes that one-egg twins are strictly identical, which in turn follows from the assumption that heredity is locked in the cellular nucleus. But Darlington (1970) refuted this latter hypothesis by finding extracellular DNA. Therefore one-egg twins, though possessing identical nuclear DNA, might differ in their non-nuclear DNA. If this difference proved to be significant, then the "identical" twins would not be strictly identical and the twin method would not be fully reliable. But this is still *sub-judice*.

Regardless of the reliability of the twin technique, the methodologically important point is that every experimental technique presupposes some hypotheses and so is just as fallible as these. (For the opposite view see, e.g. Bachrach (1962) or Rapp (1980).) In other words, experimental results are not more reliable or certain than the hypotheses underlying them, and they are not atheoretical anyway. (To be sure, the routine operations that can be performed by technicians once the original experiment has been designed do not require any knowledge of the hypotheses that went into the original experimental design. But this is beside the point: we are concerned with original research.)

Empiricism has extolled the value of the so-called *empirical basis* and its theory-free nature. There *is* no such pure theory-free empirical basis: every worthwhile experiment presupposes a number of hypotheses and, if conclusive, it confirms or undermines some hypotheses or calls for new ones. An experiment calls for a new hypothesis or theory if the existing ideas do not account for its outcome, i.e. do not explain satisfactorily how the experiment works or do not predict well its results. But of course no set of experimental results, no matter how numerous, precise and accurate, will suggest unambiguously a theory. For example, "Thousands of learning experiments with rats have not yielded a set of principles from which the performance of rats can be derived satisfactorily, although a great deal of progress has been made in the 75 years or so since the first rat was introduced into the first maze" (Bitterman, 1975).

Experiment is so essential to science and technology that it is sometimes believed to be all-important, and consequently laboratory work is regarded as a goal in itself. This is a one-sided view because it ignores the many hypotheses involved in every experiment, as well as the impact that experimental findings may have on existing hypotheses or on the construction of new ones. Like other methodological errors, this one has serious negative effects on inquiry. Here are some: (*a*) concentrating interest on variables, details, or components that are easily controllable or measurable; (*b*) focusing on instruments and techniques (e.g. electron microscopy and electrophysiology) to the detriment of the object of study; (*c*) neglecting naturalistic situations and disregarding the environment (particularly in genetics and psychology) as well as experimental "artifacts"; (*d*) underrating theory—which suggests, guides, explains, and often replaces experiment; (*e*) adopting an anthropomorphic view according to which no event happens, or at least matters, unless it is elicited and measured by some experimentalist (as is the case with the Copenhagen interpretation of quantum mechanics). Such negative effects of the exclusive concentration on laboratory work are visible in all the sciences, particularly the less developed natural sciences. The only effective antidote is a strong dose of a mixture of (good) theory and (good) philosophy.

Having warned against exaggerating the importance of experiment, let us now warn against any attempt at curtailing its scope. To be sure, some areas of inquiry are still closed to the experimental method—e.g. astrophysics and ethics. However, the experimental method has made decisive inroads in a number of areas that used to be regarded as impregnable to it. *Example 1*: Experimental astronomy started with the first Sputnik (1957). *Example 2*: Experimental geology, conducted on small scale models, has been in existence for decades. *Example 3*: Experimental palaeontology is on the way. Thus, by working on models of the dermal bony plates along the back and tail of *Stegosaurus*, it has confirmed the hypothesis that they served a thermoregulatory function (Farlow *et al.*, 1976). *Example 4*: Experimental archeology, more than a century old, has succeeded in explaining the use and efficiency of a number of artifacts and techniques, from food storage to earthwork to stone working and pottery manufacture (e.g. Coles, 1973). *Example 5*: Experimental sociology, tried sporadically and on a small scale several times during this century, was launched on a large scale in 1968 with the New Jersey income maintenance experiment. Moral: Never prophesy that a given field of factual inquiry will remain forever beyond experimental control.

4. CONCLUDING REMARKS

Whereas self-evidence is the gate to self-deception, hard evidence is the seal of adequacy—be this truth or efficiency. Evidence for or against data, hypotheses or theories, techniques, rules or plans, can be conceptual (in particular theoretical) or empirical. In other words, an idea is supported or undermined by other ideas or by empirical tests, from observation and measurement to experiment and action. And the evidence may be intradisciplinary or interdisciplinary (i.e. coming from another field of inquiry).

All empirical tests involve hypotheses and often call for further hypotheses or even entire theories. The hypotheses involved in the design and interpretation of tests are not always explicit: sometimes they are just biases. (Thus different field linguists studying the same population are bound to detect some different phonemes, because each expert expects certain sounds—which he duly records—while he does not expect others—which he may fail to record.) In science and technology there is no purely empirical basis on which to build theories, and no theories that could not possibly be affected, one way or another, by empirical tests. This is not to devaluate data but to remind ourselves that, unlike the anecdotal data of ordinary life, those of science and technology are "contaminated" by theory. Nor is it to revile speculation but to remind ourselves that all speculation must eventually be confronted with hard facts—for facts are, after all, the subject matter of science and technology.

Not all data are equally valid. In fact it has become customary, particularly in the social sciences, to class them into *hard*, *semihard*, and *soft*. Hard data are those provided by measurement or experiment on things of known composition or environment. In general they are repeatable (to within experimental error). Semihard data result from the observation of things of poorly known composition and environment. In general they are nonrepeatable. Statistical data, particularly in the social sciences, are of this kind. Finally, the soft data are occasional or anecdotal: they come from casual observation or hearsay, and they are nonrepeatable. In general, the more and better theory has gone into the production of a datum, the harder it is.

The above classification concerns the source or mode of production of data, not their intrinsic worth (truth and usefulness). In fact a hard datum can be false, or at least imprecise, or it can be useless. (This does not mean that there are no hard *facts*, but instead that our knowledge of them, even when obtained the hard way, may turn out to be false.) And although anecdotal data constitute no evidence, they are occasionally more valuable than many a trivial measurement or experimental result. For example, the observation of caged and free-roaming apes has triggered or supported much important scientific work in antropology, psychology and sociology. The quality of the data source (hard, semihard or soft) is important but not more so than the output: we judge the reliability of data sources by their performance, not conversely. (For the contrary thesis, that the plausibility of propositions depends upon the reliability of their sources, see Rescher (1976).)

The best data are those that constitute evidence for or against some hypothesis or theory and those which are precise, accurate, and varied. We will not insist on the importance of the theoretical relevance of data: suffice it to note that nobody collects data for their own sake, but always to some purpose, such as trying to discern a pattern or to check a hunch. As for repeatability, essential to minimize error, it is not always obtainable, particularly in the historical sciences. Variety, on the other hand, is in principle always obtainable because every property is lawfully related to some other properties. And it is desirable to minimize error. Thus palaeobiologists do not rely exclusively on fossils and the theory of evolution: they also make use of geological data and of information about living organisms.

The standards or criteria for evidence have become increasingly exacting in all fields of inquiry, from science and technology to the humanities and the law, since the beginning of the Modern Era. Thus hearsay and circumstantial evidence are no longer accepted, either in the laboratory or in the court of law, except as hints pointing to further search for solid evidence. Nor, of course, is a mere statement of belief or authority accepted: instead, evidence for truth or efficiency, as the case may be, is required. This dramatic progress in the standards of accuracy can be explained not only as an aspect of intellectual advancement but also by external motives. As science and technology increase in cultural and economic value, it becomes increasingly important to know to what extent they work and in which respects they are still deficient. And as the struggle for economic, political and cultural rights proceeds, the law is required to observe them more faithfully, to which end it must substitute objective evidence for denunciation and confession.

The data generated by observation, measurement or experiment may discharge several functions. They may pose problems and illustrate hypotheses. They may call for, and partially suggest, new hypotheses or theories. They may "activate" known hypotheses or theories by filling some blanks in them (i.e. assigning values to constants or parameters). Or the data may be used to assess the worth of our ideas: the truth of hypotheses and theories, the precision and accuracy of techniques and instruments, or the efficiency of rules, proposals, and artifacts. Let us study next the latter operation, namely evaluation.

CHAPTER 12

EVALUATING

All cognitive activities are motivated. We seek to find, think to solve problems, plan to attain goals, talk and listen to communicate, and so on. And we evaluate all the conscious activities, whether purely cognitive or using some knowledge. We evaluate problems and their solutions, propositions and proposals, theories and designs, methods and artifacts. We evaluate some of them for truth and others for usefulness: these are the supreme values in matters of cognition. Any other values, such as simplicity and beauty, are subsidiary with regard to knowledge. To be sure, a false hypothesis and a useless machine may be valuable in some regard—e.g. heuristically in the former case and commercially in the latter—but ultimately they are to be discarded.

Those two main values, truth and usefulness, are quite distinct: so much so that certain false propositions are valuable (for evil purposes) whereas some true ones have no practical use. This does not entail that truth and usefulness are always unrelated. Modern technology involves both, for it employs theories that have passed truth tests, and does so in order to guide the production or maintenance of (supposedly) useful things. What the distinction does entail is that truth is no substitute for usefulness or conversely, whence the search for truth is not the same as the search for usefulness. In particular, science is not identical with technology. In a way the latter is more demanding than the former, for it uses and produces both truth and usefulness.

Although truth and usefulness are the supreme values of cognition, they are quite elusive: they are difficult to define and recognize. To be sure, there are precise definitions of logical and mathematical truth, but they are inapplicable to factual propositions. And anyone can draw a list of useful items, but few can say what value is. We shall outline a program for a correspondence theory of truth, and another for a theory of value (Section 1.)

Being seldom manifest, truth and usefulness are often hard to pinpoint. As in the case of other unobservables, we need indicators. I submit that there are both empirical and conceptual indicators of truth and usefulness. The foremost empirical indicator of truth is confirmation, and that of usefulness efficiency (Section 2). And the main conceptual indicator of either is compatibility with our background knowledge (Section 3). We come up then with a battery of truth indicators and another of usefulness indicators. Each battery allows one to assign, at least temporarily, degrees of truth or of usefulness, as the case may be. I wrote 'assign' not 'disclose' or 'discover', because—and this is a central thesis of this chapter—neither truth nor usefulness are intrinsic and eternal properties: they are contextual and changeable. As von Uexküll (1928, p. 1) said, scientific truth is today's error, and the progress of science is a process of moving from gross to fine error. Likewise we may say that usefulness is today's waste, and the progress of technology is or ought to be one of decreasing waste.

1. VALUES

1.1. *Truth*

We all value adequate knowledge: adequate propositions (in particular data and hypotheses) and adequate proposals (in particular rules and plans). Our notion of adequacy embraces two concepts, those of truth and usefulness. An adequate proposition is said to be true (to a considerable extent), whereas an adequate proposal is said to be useful (to a considerable extent). Thus the proposition that crème caramel is nourishing, is true, whereas a recipe for baking a good crème caramel is useful. In this section we shall study truth, leaving usefulness for the next.

Let us begin by recalling briefly some of the ideas on truth discussed at length in Vol. 2, Ch. 8; after that we shall have a new look at the problem. To begin with we stipulate that truth (total or partial) is a property of propositions that have passed certain tests. This statement raises immediately two questions: Why should propositions, rather than sentences, be the truth bearers, and why should propositions lack a truth value unless they are first tested? The answer to the fist question is this. Firstly, not all sentences are meaningful, hence not all of them could be assigned a truth value. For example, 'This book squeezes my anvil', though grammatically correct, has no cognitive content and therefore does not designate a proposition. Second, the language in which a proposition is expressed is important only for purposes of communication. Thus the sentences 'Russell was a famous thinker' and 'Russel fut un penseur célèbre', as well as their renderings in American Sign Language, express the same (true) proposition.

As for our view that propositions are not born with truth values but are assigned truth values on the strength of tests, it is at variance with tradition. According to this tradition, which probably originated with Plato, truth (or falsity) is possessed by a proposition whether anyone knows it or not. This view is still found, though usually unexamined, in much of contemporary epistemology. In particular epistemic logicians usually equate "s knows p" with "s knows that p is true". Yet there are two different concepts here: subject s may know or understand p without having the slightest idea about the truth value of p. For example, a theorist who conjectures or infers a hypothesis p will not know whether p is factually true until p has been subjected to empirical tests—and even then some uncertainty may remain. In short, a proposition p must be distinguished from the metastatement "p is true (or false)", which in turn can only be a result of an investigation of p.

We adopt then the constructivist thesis that propositions, and in general constructs, exist (formally) if, and only if, they are knowable. In particular, truth values are acquired or lost as a result of tests for truth. This does not prevent us from *feigning*, for the purposes of logical or mathematical work, that there are constructs, in particular propositions, detached from the material world and, in particular, from thinking brains (cf. Bunge, 1981). We adopt this as a useful fiction and as tantamount to saying that, for certain purposes—such as logical analysis and mathematical calculation—it does not matter who thought up the propositions concerned or under what circumstances he did so. Pretending that propositions exist in this manner is radically different from claiming that they exist in and by themselves, the way Plato, Bolzano, Hegel, and Popper held. Our view is then compatible with the naturalist ontology formulated in Vols. 3 and 4.

Our constructivist thesis is superficially similar to that of operationism, according to which physical objects have no properties as long as these are not measured. The analogy is superficial, for physical objects happen to exist by themselves, and every existent has some properties. On the other hand constructs must be constructed by someone in order to exist. It is we who, by fashioning them—i.e. thinking them up—endow them with properties, in particular truth values—or abstain from doing so. Table 12.1 summarizes our view so far.

The kind of test that may determine the truth value of a proposition depends on the kind of proposition. Thus the (formal) truth of the Pythagorean theorem is established by deduction within Euclidean geometry, whereas its factual adequacy is investigated by measure-

TABLE 12.1
Truth not inborn but acquired or lost

Methodological viewpoint	*Semantical viewpoint*	*Pragmatic viewpoint*
p has not yet been investigated, or it has been studied and no conclusive results have been obtained.	*p* has no truth value.	Suspend judgment about the worth of *p*: neither believe nor disbelieve *p*.
p has been investigated and found coherent with both background knowledge and new data.	*p* is true (to some extent).	Accept *p pro tempore*.
p has been investigated and found inconsistent with background knowledge and new data.	*p* is false (to some extent).	Reject *p pro tempore*.

ment—which shows it to be approximately true of physical space. There are then, as Leibniz argued, truths (and falsities) *of reason*, and truths (and falsities) *of fact*. We count the logical, mathematical and semantical among the former, and the ordinary, scientific, technological, ontological and epistemological among the latter. A truth of reason is recognized by its coherence with some body of knowledge, i.e. by purely conceptual operations. On the other hand a truth of fact is recognized not only by its matching a body of (factual) knowledge but also by its passing certain empirical tests showing its adequacy to concrete or material reality. See Figure 12.1.

Because formal and factual truth are predicated of propositions of radically different kinds, and on the strength of operations of diverse types, each of them requires its own theory. That is, we need a *coherence* theory to elucidate the notion of formal truth, and a *correspondence* theory to explicate that of factual truth. The claim that a single theory of truth can perform both tasks is consistent only with either an idealist ontology or a radical empiricist epistemology. In the former case reality, or the actual world, is just one more model of some abstract structure. In the latter case all knowledge, whether formal or factual, is acquired only through sense experience: the task of reason would be only that of ordering or systematizing such knowledge, never generating or validating it. Since we accept neither an idealist ontology (cf. Vols. 3 and 4) nor an empiricist

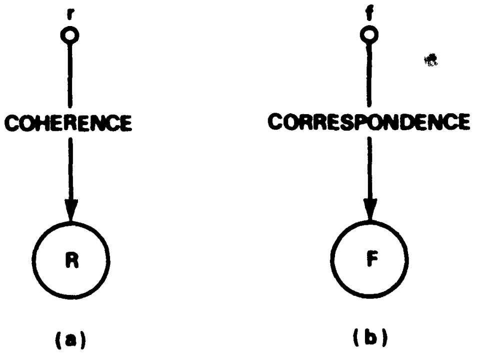


Fig. 12.1. Two kinds of truth. (*a*) Formal truth: proposition *r* coheres with body *R* of propositions. (*b*) Factual truth: proposition *f* corresponds to facts *F*. Since *R* and *F* are disjoint, the relations of coherence and correspondence are different.

epistemology (as is clear from this volume), we reject the monistic view of truth, which conflates formal and factual truth, and we undertake to find one theory for each kind of truth (or falsity).

Now, the coherence theory of truth is available: it is part of model theory, or the semantics of logic and mathematics, and it is usually included in logic. On the other hand the correspondence theory of truth is so far little more than a program. We shall not be concerned here with formal truth and its theory, i.e. model theory, of which more in Vol. 2, Ch. 8, and Vol. 6. We shall grapple instead with the idea of truth as correspondence with fact. Our investigation will proceed in two stages. In the first we shall study what the correspondence consists in; in the second we shall investigate the actual process of truth value assignments, or the problem of truth criteria.

Our first problem, then, is that of clarifying the sense of 'correspondence' or 'adequacy' in the scholastic formula *Veritas est adaequatio rei et intellectus* (or *intellectus ad rem* in a more realistic version). How can a proposition, which is a construct, be adequate to (or match or fit in with) a fact in the world of things? One answer is that there is an isomorphism between a proposition and the piece of reality it refers to: this idea has been upheld by naive realists and Gestalt psychologists. However, 'isomorphism' is here as little a technical term as 'adequacy' in the scholastic formula. Indeed, isomorphism can exist only between sets, and neither a proposition nor the concrete thing it refers to are sets. This difficulty is seemingly solved by postulating that sets of propositions, such as theories, can correspond to collections of facts. But then truth would be a collective or emergent property possessed by a theory, not by its constituents. Besides, sets of facts have rarely a definite structure. Systems do have definite structures, but not every thing is a system. Besides, it is easy to exhibit systems that are not isomorphic with their theories—e.g. a system composed of two particles gravitating around each other is not isomorphic to any theoretical model of it, which model contains infinitely many propositions.

We seem to be up against a Gordian knot, so let us cut it. Let us not pretend any longer that the truth relation holds between propositions and facts. Let us admit instead that propositions can match other propositions, and facts other facts. And let us assume that the correspondence we are seeking is one between mental facts of a certain kind and further facts, whether mental or not. Further, in line with psychobiology we identify mental facts with brain processes occurring in plastic neural systems (Ch. 1). This strategy will allow us to characterize (though not define in general terms) the concept of true partial knowledge of fact; once in possession of this notion we shall proceed to define the truth of a proposition.

Consider a thing a internal or external to an animal b endowed with a brain capable of learning. Call e an event (change) in thing a, and e^* the corresponding perceptual or conceptual representation of e in the brain of b. Then we say that b has gained *true partial knowledge of fact* e if, and only if, e^* is identical with the perception or conception of e as a change in thing a (rather than as a nonchange, or as a change in some other thing). The true partial knowledge that b has acquired of event e is the neural event e^*; and the *correspondence* involved in it is the relation between the events e and e^*.

Consider next the case where the object of knowledge is a binary relation between events, such as the temporal succession of two flashes of light. Let now e_1 and e_2 denote changes occurring in a thing a internal or external to an animal b capable of learning, and call e_1^* and e_2^* the perceptions or conceptions that b forms of e_1 and e_2 respectively. Further, let e_1 be R-related to e_2, i.e. Re_1e_2. We stipulate that b has *true partial knowledge of the fact that* Re_1e_2 if, and only if, (a) b recognizes e_1 and e_2 as changes in thing a (rather than either as nonchanges or as changes in something else), and (b) b perceives or conceives of e_1 and e_2 as being R-related, i.e. if $e_1{}^*$ and e_2^* are R-related. The true partial knowledge that b has gained of Re_1e_2 is $Re_1^*e_2^*$, and the correspondence involved is the relation between the event pairs $\langle e_1, e_2\rangle$ and $\langle e_1^*, e_2^*\rangle$. In general, the brain correlate of Re_1e_2 will be $R^*e_1^*e_2^*$, where R^* is similar to but not identical with R.

Consideration of the above two cases, of the single event and the pair of events, suffices for our present purposes. Let us now move from thoughts (brain processes) to propositions (constructs). Recall that we have defined a construct as an equivalence class of thoughts (Ch. 1, Section 2.1.) Consider the class of thoughts e^* elicited in animals of a given species by event e, and form the subset of e^*s that happen to constitute true (though partial) knowledge of e. Call this class $[e^*]$. No two members of $[e^*]$ may be identical—for they are thoughts of a given animal at different times or thoughts of different animals, and in either case they differ in some respect or other. However, they are all *equivalent* in that every one of them constitutes partial true knowledge of e; i.e. for every member e^* of $[e^*]$, e^* happens if, and only if, e is the case. We identify the proposition $p = \ulcorner e$ is the case$\urcorner$ with that equivalence class of thoughts, i.e. we set $p = [e^*]$. And we stipulate that *p is true if and only if e happens.* Likewise in the case of pairs of R-related events discussed above we form the equivalence class of thoughts $[\langle e_1^*, e_2^* \rangle]$ and identify it with the proposition $p = \ulcorner Re_1e_2 \urcorner$, stipulating that *$Re_1e_2$ is true just in case $R^*e_1^*e_2^*$ is the case.* In short, the correspondence relation that holds between a mental fact and some other (mental or nonmental) fact carries over to propositions in relation with facts. In other words, we have analyzed the fact-proposition correspondence into the relations shown in the following diagram:

$$\begin{tikzcd}
\textit{Facts } E \arrow[r] \arrow[dr] & \textit{Thoughts } E^* \arrow[d] \\
 & \textit{Equivalence class } E^*/\sim \textit{ of thoughts} \\
 & \textit{(proposition)}
\end{tikzcd}$$

In sum, we have identified factual truth with adequate knowledge of fact: we say that the proposition "e is the case" is true if, and only if, some animal has adequate knowledge of e. On this view truth and falsity are *primarily* properties of perceptions and conceptions, and only *secondarily* (vicariously) properties of those equivalence classes of thoughts we call 'propositions'. Hence a faithful perception or representation of a human figure can be said to be true, whereas a hallucination can be said to be a false perception. Moreover we identify the organ of truth and falsity as the plastic brain. It is a problem for neuropsychological research to try and localize, within the human brain, the neural systems that attribute truth and falsity.

Note that we have dealt only with simple (atomic) propositions, not with complex (molecular) ones such as disjunctions and generalizations. It might be thought that the latter can always be built out of simple ones, so that the truth values of the former can be evaluated on the basis of the truth values of the simple propositions together with a set of semantical truth conditions. This belief is false, as shown by any of the differential equations occurring in theoretical science. Thus a basic equation of Maxwell's electromagnetic theory reads: "For every electromagnetic field and the current that generates it, and every point in spacetime, the d'Alembertian of the vector potential of the field is proportional to the current density". This formula cannot be construed as an infinite conjunction of simple statements, one for each field and its corresponding source, and for each point of spacetime. Indeed the points of spacetime constitute a continuum, and logical conjunction is not defined for a nondenumerable set of propositions. The semantic truth conditions fail us here and we must resort to entirely different truth conditions, supplied by experimental physics and bearing on the testable consequences of the axiom in question.

Note also that we have dealt only with individual propositions not with hypothetico-deductive systems. One reason for this restriction is that an entire theory is unthinkable for it contains infinitely many propositions. We can think only of a few statements (postulates, definitions or theorems) of any given theory. When we think of a theory as a whole, in point of fact we select a handful of typical formulas of it. Therefore attributing truth or falsity to a theory is making an inductive leap on the basis of some evidence. (More in Section 2.2.)

It will be noted also that in the above discussion we spoke of *partial* knowledge of fact. Equivalently: factual propositions represent only some (not all) features of things, or they are not totally true but rather half true. Total truth and falsity, which are the only admissible ones in logic and pure mathematics, are hard to come by elsewhere. In applied mathematics, science, technology and the humanities, no less than in ordinary life, most statements are at best approximately true. In most cases we must settle for half truths in the hope of replacing some of them with three-quarter truths. Think of the rounding off of figures, the method of successive approximations, the errors of measurement, and the simplifications made in any mathematical model. Everywhere except in logic and in pure mathematics we deal with approximations and do our best to improve them. Therefore it behoves the philosopher to study error and its complement, partial truth. See Appendix 3.

Twelve centuries before Descartes' *Cogito, ergo sum*, Augustine wrote: "If I err, I am. For he that has no being cannot err, and therefore mine error proves my being. Which being so, how can I err in believing in my being?"

(*The City of God* Bk. XI, Ch. XXVI). Error is our constant companion, hence we should learn to live with it. We do so when learning from error and learning to correct it—which, by the way, is less frequent than learning from success. All observation, measurement or experiment is subject to error, and so is every nondeductive—in particular inductive and analogical—reasoning. However, without such operations we would hardly gain new knowledge. The rule therefore is not "Avoid error", good only for Superman. It is, rather, "Be not afraid of error but of refraining from exploring for fear of erring, and of being unwilling to correct error and learn from it". Moreover we should cultivate controlled error, or what passes for such, by questioning once in a while our most cherished beliefs and modes of thinking, to see whether they need alteration or, on the contrary, can be trusted for another while.

As there are two kinds of truth so there are two kinds of error: of reason and of fact. An *error of reason* is a mismatch among thoughts (or the corresponding propositions), whereas an *error of fact* is a mismatch between thoughts (or the corresponding propositions) and their factual referents. (However, both are brain processes and in this sense factual.) Each kind of error falls in turn into one of the following categories: of *omission* (overlooking propositions or facts) and of *commission* (misconception, misperception, or misdeed).

The most common errors, whether of omission or of commission, are of the following kinds:

(*a*) *erotetic*: asking the wrong question;
(*b*) *postulational*: assuming the wrong premises, be these presuppositions or explicit hypotheses or data;
(*c*) *dialectical*: advancing the wrong argument;
(*d*) *methodical*: choosing the wrong method or technique;
(*e*) *design*: designing the wrong operation, measurement, experiment, plan, or artifact;
(*f*) *computation*: making calculation mistakes;
(*g*) *sampling*: inferring from unrepresentative samples;
(*h*) *axiological*: wrong evaluation of, e.g. a proposition or a goal.

We say that an error is *wrong-headed* if it combines two or more errors of the above kind—e.g. the wrong problem with the wrong approach. (*Example*: Asking at what time the soul enters the human embryo, and looking for the answer in a theological treatise.) Characteristically, ordinary knowledge is plagued with errors of all the above kinds, often to the point of being wrong-headed. Worse, ordinary knowledge errors tend to be self-perpetuating (in popular myth) rather than self-correcting (as in science and technology). In particular, every time we announce a generalization from a hasty personal sample we commit a sampling error, often compounded by the contrary statement suggested by an equally nonrepresentative sample.

Design errors deserve a special comment. In measurement or experiment they are called *constant, systematic*, or simply *biases*. They are mistakes in the design of the operation, e.g. lack of adequate controls, or electrical leaks. Once detected they may be corrected, often completely. Such constant errors must not be confused with random errors, such as typing errors and the fluctuations around the average of a set of measured values. These other errors are really random: they have a regular statistical distribution (usually bell-shaped or gaussian). Although they can often be reduced, they can never be completely eliminated because they are rooted in objective randomness—e.g. the random thermal motions in both the measured object and the measuring instrument. (For this reason they should not be called *random errors* but perhaps *random scatters*.) Since chance is universal, random errors can be studied regardless of the nature of the measurement or experimental set-up. In fact they are studied by the mathematical theory of errors of observation, an application of the calculus of probability. On the other hand, since the constant (systematic) errors depend upon the nature of the components of the experimental set-up or its manipulation, there can be no general theory of constant (or design) errors.

Luckily we possess error detection mechanisms and have invented error correction procedures; moreover, both are perfectible with learning. The human brain seems to be equipped with error detecting and correcting systems—sensory, motor, and conceptual. The efficiency of such systems improves with (successful) practice and deteriorates in pathological conditions. For example, under the action of neurotropic drugs, or in the presence of brain disease or injury, such error detectors fail, and consequently the performance of mental activity becomes suboptimal; the electrical stimulation of such detectors affects them likewise (Bechtereva, 1978, pp. 134ff.)

Wiener (1948) conjectured that every error-correcting mechanism is cybernetic. To understand this recall that every error is a discrepancy or deviation, e.g. from the truth, or from some desired or normal value. Thus, in the case of a numerical variable, if the actual value is *a* and its correct (or optimal) value c, then the error in that variable is the absolute value of their difference, i.e. $\varepsilon = |a - c|$. If this discrepancy in the output of the system concerned overtakes a tolerable threshold, the control system is activated, and adjusts the input so that the error will decrease. This cycle is repeated until the (intolerable) error vanishes. We have become accustomed to thinking that such negative feedback mechanisms are common to inquiring systems and artificial control systems.

No doubt, errors in the performance of sensory-motor tasks and elementary intellectual operations are likely to be corrected (if detected) in this manner, i.e. by activating feedback loops in existing neural systems. However, the recognition and correction of bulkier errors, such as choosing the wrong problem, the wrong approach, or the wrong world view, may require far more than an existing feedback mechanism. Such errors may call for a radical reshuffling of neural connections—i.e. a self-reorganization—and the corresponding emergence of new conceptual systems.

When rendered conscious, the error correcting processes take the form of *methods* such as the dialectical method, the hypothetico-deductive method, the method of successive approximations, the statistical method, and the experimental method. One uses Socrates' (not Hegel's) *dialectical method* when investigating a set of premises, such as hypotheses and data. If they turn out to be utterly irrelevant or false, one discards them; but if they are relevant and one finds a grain of truth in them, one attempts to refine them, render them consistent with the antecedent knowledge, or adjust some of the latter to the new finding. (This procedure is not dialectical in Hegel's sense because it does not involve concepts that "contain their own negation"—whatever this may mean—and does not advance from contradiction to contradiction but from ignorance to knowledge, from falsity to approximate truth, from stray bits to systems, or from shallowness to depth.) The *hypothetico-deductive* method, or method of hypotheses, is a particular case of the dialectical method: it consists in advancing hypotheses, investigating their consequences, and adjusting or eliminating the former if such consequences turn out to be false.

The *method of successive approximations* is another particular version of the dialectical method. In mathematics it usually works as follows. One poses a certain problem and tries to solve it. If the problem cannot be solved exactly (or in closed form) one obtains a first approximation, which serves as the basis for constructing a second order approximation, and so forth, until a sufficient approximation, or even an exact solution, is reached. In factual science and technology the method of successive approximations works similarly but not quite, for in this case there are no algorithms for advancing securely from one hypothesis to the next: one proceeds in zigzag until hitting on an acceptable hypothesis—or on none. However, one does follow a prescribed sequence of steps such as: Initial data–hypothesis 1–data 2–hypothesis 2–. . . –hypothesis n, or some other alternate chain of conjecture and evidence.

The *statistical method* is a particular case of the method of successive approximations, in which the hypothesis, which concerns an unknown population, is suggested or checked by investigating a random sample of it, and rechecked by resampling. Finally, the *experimental method* is the same thing, except that experiment is explicitly involved. All five methods are error correcting (or rather minimizing) procedures. And the last three—the method of successive approximations, the statistical and the experimental method—presuppose that truth is usually partial rather than total.

Error is the complement of truth: the smaller the former the greater the latter. In symbols, $V(p) + \delta = 1$, where $V(p)$ is the truth value of proposition p, and δ the relative discrepancy or error of p relative to some other proposition taken as a base line. The latter may be a datum if we are checking a hypothesis, or a law statement if we are checking a datum. In writing the above formula we have assumed that truth values are real numbers comprised between 0 (utter falsity) and 1 (complete truth). That is, we have assumed that V is a function from a set P of propositions (those that can be assigned a truth value) to the unit interval [0, 1] in the real line. We have also assumed that the truth value of a proposition is relative to some other proposition, which is not questioned when evaluating the former although it may of course be questioned in a different context.

TABLE 12.2
Truth value of proposition p assuming base line q is true, i.e. $V(p) = 1 - \delta(p,q)$

Discrepancy $\delta(p,q)$	*Truth value* $V(p)$	*Vernacular*
0.0	1.0	p is true relative to q (rel. q)
$0.0 < \delta \leq 0.2$	$1.0 > V(p) \geq 0.8$	p is approximately true (rel. q)
$0.2 < \delta \leq 0.4$	$0.8 > V(p) \geq 0.6$	p is half true (rel. q)
$0.6 < \delta \leq 0.8$	$0.4 > V(p) \geq 0.2$	p is three quarters false (rel. q)
$0.8 < \delta \leq 1.0$	$0.2 > V(p) \geq 0.0$	p is nearly false (rel. q)
1.0	0.0	p is false (rel. q)

Therefore our initial formula should be generalized to read

$$V(p) = V(q) - \delta(p, q).$$

Table 12.2 exhibits a few typical cases.

The discrepancy $\delta(p,q)$ may be equated to the *reliability index*, a widely used measure of the theory-data discrepancy. This index is defined as follows. Let p and q be propositions of the form "The value of property M for thing a is r", or "$M(a) = r$" for short, where r is a real number. If both p and q refer to the same thing, and q is adopted as the base line, then the discrepancy is defined as

$$\delta(p,q) = \frac{|m - m'|}{|m'|},$$

where m and m' are the numerical values of the property M according to propositions p and q respectively, i.e. $p = \ulcorner M(a) = m \urcorner$, $q = \ulcorner M(a) = m' \urcorner$.

The above reliability index or discrepancy can be generalized to sets P and Q of propositions, such as a set of theoretical formulas and the corresponding set of empirical data, by summing over the corresponding values:

$$\delta(P, Q) = \frac{\sum |m - m'|}{\sum |m'|}.$$

In particular, this formula covers both the case of checking theoretical calculations against observations, and of checking observations against calculations. But it applies only to a smallish sample of all the propositions of a theory, namely those that are actually subjected to tests. Moreover, that sample is not a random one, so if the theory-data discrepancy is small (large) we cannot safely infer that the theory itself (together with the subsidiary assumptions and the data necessary to derive testable consequences) is true (false). Finally, the above formulas apply also to the confrontation of theoretical propositions with other theoretical propositions, and of empirical data with other empirical data, provided the confronted pairs of propositions share a sense and a reference. For more on the way (partial) truth values are assigned in science see Vol. 2, Ch. 8, Sections 2.2–2.4. For theories of partial truth see Appendix 3.

Two warnings are in order. One is that, in general, there is one discrepancy value for each value of the variable(s) under experimental control. Hence as the experimenter explores different parts of the theoretical curve the discrepancy may well vary, i.e. increase or decrease.

Moreover this variation is *lawless*, because there is no logical or physical necessity linking the experimental data to the theoretical predictions.

The second warning is that experimental accuracy depends not only upon the quality of the instrumental, e.g. the laboratory equipment, but also upon the mathematical form of the formulas being checked—which form is to some extent conventional. A simple example will suffice to make this point. Suppose the task is to check a theoretical formula of the form "$y = ax/(b + x)$". This task is complicated by the fact that a is the asymptotic value of y, i.e. $y(\infty)$: see Figure 12.2. To facilitate the task, one performs the simple trick of transforming the given formula into a linear equation by means of the variable changes $x = 1/u$ and $y = 1/v$, which ensues in "$v = (1/a) + (b/a)u$". Now a is the reciprocal of the ordinate value at the origin, usually an accessible number.

Inductive logicians have proposed a number of acceptance and rejection rules for propositions, that are independent of the concept of truth. Scientists appear to abide, albeit tacitly, by a different set of rules, namely

$R1$: Accept (*pro tempore*) all and only the (sufficiently) true propositions.

$R2$: Reject (*pro tempore*) all and only the negates of the (sufficiently) true propositions.

$R3$: Suspend assent or denial of all and only the propositions that you neither accept nor reject (*pro tempore*).

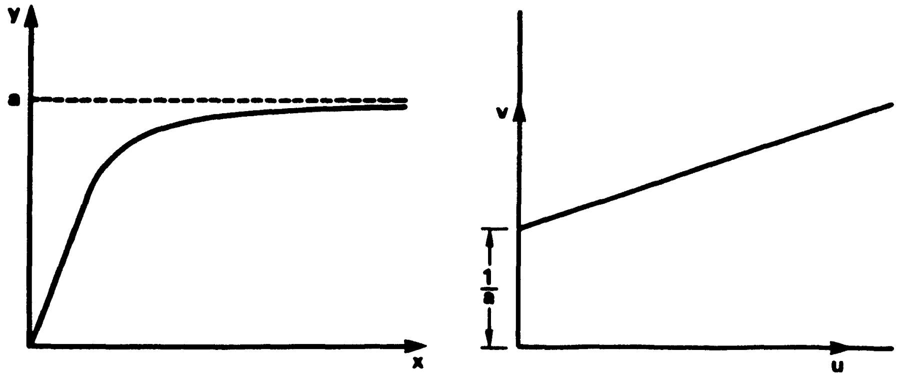
These rules can be unified and somewhat exactified by introducing three functions from the set of propositions to the set of decisions: A (acceptance), R (rejection), and S (suspension). The last two functions can be defined in terms of the first, namely thus:

$$\text{For any proposition } p,\ R(p) =_{df} A(\neg p),\ S(p) =_{df} \neg A(p) \,\&\, \neg A(\neg p).$$

The above rules reduce then to the single rule

R: *For all propositions* p, $A(p)$ *iff* $V(p) > l - b$, where b is the tolerance or maximal discrepancy allowable given the state of the art.

We close with a batch of assorted remarks. First, Wittgenstein (1953) and his followers have claimed that the concept of truth is dispensable: that every proposition of the form "p is true" can be replaced without loss with "p". This no-truth thesis about truth is logically and factually mistaken. It is a logical mistake because "p is true" is a metastatement, not a statement. (In other words, $p \neq \ulcorner V(p) = 1 \urcorner$.) The thesis (known as "the absolute theory of truth") is factually erroneous because scientists and technologists want to know whether their hypotheses are true, and this is why they check whether they "fit" the data. (The very notion of goodness of fit involves the concept of partial truth.)



(a) (b)

Fig. 12.2. Accuracy dependent upon mathematical form of hypothesis. A mere change of variables transforms the arduous task of determining the asymptotic value a into the far easier one of measuring the ordinate $1/a$ at the origin in (b).

A related doctrine is conventionalism and, in particular, fictionism, according to which the scientist does not seek truth because he investigates phenomena not reality. "He seeks system, simplicity, scope; and when satisfied on these scores he tailors truth to fit" (Goodman, 1978, p. 18). Let us gloss over the unfortunate fact that the author of this quote does not explicate the notions of simplicity and scope: suffice it to say only that, if scientists and technologists did not care for truth, and could make it up, they would not devote so much time to checking their information and testing their hypotheses for truth.

It is usually held that truth values do not really change in the course of time: what may happen is that sometimes we fail to discover their real truth values. This view presupposes that truth values are intrinsic or absolute rather than relative to some body of (empirical or theoretical) evidence. The obvious question is: How does anyone know that a given proposition has a certain truth value before it has been tested (or perhaps even formulated)? As long as no satisfactory tests have been performed we had better abstain from assigning truth values, i.e. from formulating metastatements of the form "$V(p) = v$". Factual truth values are assigned on the strength of tests and, since these are fallible, truth values are transient. This does not entail that truth, like fashion, is quite arbitrary: it does entail instead that truth, like fashion—and like everything else in matters of knowledge—is changeable.

Does the transience of truth values mean that factual knowledge is only probable in the technical sense, i.e. that truth is the same as probability? Clearly, truth differs from objective (physical) probability, or propensity.

Thus consider a thing that can be in any of n different states with equal probability $1/n$. The truth value of "The thing is in state m", where m is one of the n states, is either 0 or 1. This example suggests that there is no lawful relation between objective probability and factual truth.

What about subjective probability? Could we speak of the chance of a proposition being (totally) true, and equate that chance with the subjective probability of the proposition? This is what some philosophers have proposed. Since they take probabilities to measure intensities of beliefs, it turns out that the chance of a proposition being true equals the intensity of our belief in it. In other words, they propose the formula *p is true for subject s if, and only if, s believes strongly that p.* Thus truth would be just as subjective as belief. Consequently science and technology would have no more claim to truth than myth and religion. True, most proponents of subjective or personalist probability do not draw this consequence: they are inconsistent. But we insist on evaluating philosophical views by their fruits.

Other philosophers have proposed to substitute intersubjectivity for subjectivity, advancing the slogan *True is what the majority of researchers believe.* This, the consensus view, has been defended by Fleck (1935), Polanyi (1958), Kuhn (1962), Ziman (1968), and Rorty (1979). True, consensus is an indicator of truth; but it is not the same as truth: remember the fable of the Emperor's clothes. Factual truth is not just a matter of belief but of evidence. So, the above slogan would be adequate if in turn one could prove that in all cases the majority of researchers believe only the propositions supported by plenty of evidence. But this is a matter for social psychology. In epistemology we are interested in finding out whether or not there can be such thing as objective truth estimated on the strength of tests and quite independently of the strength of the beliefs of individuals or even entire communities. This is why we care for confirmation (Section 2.1).

1.2. *Usefulness*

An item of knowledge, such as a theory or a plan, can be valuable cognitively, practically, or both. Obviously we prefer our knowledge items to be valuable in at least one way. And it is also clear that this desideratum is not satisfied by every cognitive item. Thus counting the number of pebbles on a given beach is likely to be just as worthless as devising a technique for catching ghosts. In sum, not all things are worth knowing—just as not all actions are worth performing.

Cognitive and practical value are deceivingly transparent. Actually they are rather difficult to define and attribute. Let us start by examining the general notion of value, which will be studied in detail in Vol. 7. There is, first, the naive notion of *absolute* value, such as it occurs in the question 'What is the value of this datum (or that proposal)?' This notion is imprecise, for every time we assign a value to an object we do so in some respect or other: an item may be good for something but worthless or even disvaluable for something else.

The notion we want is that of *relative* value, or value in some respect, such as it occurs in the statement that cosmology is cognitively valuable though practically worthless. We are often in a position to make reasonable qualitative value judgements such as "The value of item x in respect R is great (or small or nil or negative)". And we can analyze such a global value statement in some such way as this: "The R-value of x for thing (e.g. person) a in circumstance b and in view of desideratum c is great (or small or nil or negative)" (Bunge, 1962b).

An object that is valuable may be so in itself or as a means for attaining some other object possessing a value in itself. We speak accordingly of *intrinsic* and of *instrumental* value as two varieties of relative value. (Intrinsic value must not be mistaken for absolute value: whatever is valuable is so in some respect or other—in particular as a means or as an end.) If we could quantitate value we would say that the total value of an object equals the sum of its intrinsic and instrumental values.

(It may be possible to quantitate value, i.e. to assign a definite number to the value of every object of some kind in some respect. I.e. one might try to postulate that, for a certain kind K of objects, in a certain respect R, there exists a function V_R from K into the double unit real interval $[-1, 1]$. Hence one could write, e.g. $V_R(x) > 0$ for "x is valuable in the respect R", $V_R(x) = 0$ for "x is worthless in the respect R", and $V_R(x) < 0$ for "x is disvaluable in the respect R". This possibility will be explored in Vol. 8.)

A realistic way of looking at valuation is to do so in terms of the *relative and comparative* notion of value, occurring e.g. in the question 'Which of the items x and y is the more valuable in the respect R?' Indeed it is often possible to rank objects of a given kind (e.g. hypotheses) in a certain respect (e.g. testability), much in the way we grade students with respect to originality and industry. We shall assume that, at least in principle, every homogeneous set K of cognitive items can be ordered in a given respect, such as truth or practical usefulness. (The homogeneity condition means that we compare hypotheses to hypotheses, proposals to proposals, and so on.) In other words we postulate that $\langle K, \geq_R \rangle$, where K is a homogeneous set of cognitive items, is a partially ordered set, so that, for any x and y in K, $x \geq_R y$ or $y \geq_R x$. (I.e., $\geq_R$ is assumed to be a simple ordering of K.)

We are now in a better position to clarify the notions of cognitive and practical value. We stipulate that an item of knowledge is *intrinsically valuable* if it is, or is likely to be, a (sufficiently) true proposition. In other words, intrinsic value is, in matters of knowledge, identical to truth and, like truth, it comes in degrees. Secondly, we stipulate that an object is *instrumentally valuable for inquiry*—or *instrumental for inquiry*—if it is an epistemic operation—such as questioning or measuring—or part of one, likely to yield an intrinsically valuable (i.e. sufficiently true) piece of knowledge. Thirdly, we stipulate that a knowledge item is *instrumentally valuable for action*—or *useful* for short—if it is, or is likely to be, a means for attaining a practical goal (whether or not goal itself is valuable in a given respect). Note the cautious clause "is likely to be" in all three definitions: it is a caution not to write off a piece of information, or an epistemic operation, just because it looks worthless at first blush. It may well be that, on a more careful investigation, or under new circumstances, it turns out to be valuable.

From the above definitions it follows that truth, though always (intrinsically) valuable, may not be instrumentally valuable, i.e. may lead nowhere and, in particular, it may not be practically valuable, i.e. useful. It also follows that, for an object to be practically valuable (useful), it need not be intrinsically valuable. In particular, untruths and even lies may be useful for certain purposes—particularly though not exclusively evil ones. Finally, the cautionary "is likely to be" reminds us that, even though we evaluate every cognitive item, we should not regard such valuations as certain and definitive. To proceed otherwise is to be short-sighted.

A couple of examples will help grasp the above notions. Unveiling the stratigraphy of a given site is intrinsically valuable, for it contributes to our knowledge of our planet. And it may also be instrumentally valuable, for both cognitive and practical purposes. Indeed, knowing the kinds of rock layers and their order is instrumental to dating whatever fossils or tools may be found in the site, and it is useful to identify minerals or search for them. In other words, geology is intrinsically valuable (as a science in itself), it is valuable to other fields of inquiry, and it is also useful to the economy. Our second example is epistemology: it is intrinsically valuable for contributing to our knowledge about knowledge. But epistemology is also instrumentally valuable for inquiry, if only for clarifying certain key methodological terms such as those of taxonomy and theory. Finally, epistemology can be useful in designing science and technology policies, e.g. by elucidating the very concepts of basic research, applied research, technology, and technical service (such as the practice of engineering or medicine), and by showing that, on the whole, it pays to support the former in order to make progress in the latter.

2. EMPIRICAL VALUE INDICATORS

2.1. *Confirmation*

Truth is not manifest but there are truth symptoms or indicators. One of them is empirical confirmation or corroboration. We shall deal here with the empirical confirmation of hypotheses and theories. To begin with, we stipulate that a hypothesis p is *empirically confirmed* (or *supported*) by a datum e if and only if (*a*) e is empirical evidence for or against p (Ch. 11, Section 1.2), and (*b*) e is roughly as true as p or some logical consequence of p.

The concept of empirical confirmation can be considerably exactified in case both the proposition and the datum in question are quantitative statements, such as

p: The theoretical or calculated value of property M of object x in state y equals t. Briefly, $M_\vartheta(x,y) = t$.

e: The measured value of property M of object x in state y equals u plus or minus the relative random error σ. I.e., $M_\varepsilon(x,y) = u \pm \sigma$.

In this case we stipulate that p and e are *empirically equivalent to within the error* σ if and only if the difference between the theoretical value t and the average u of the measured values is smaller than the relative random error, i.e.

$$p \sim_\sigma e =_{df} |t - u| \leq \sigma.$$

Obviously, p and e are empirically *inequivalent* with respect to the error σ if and only if they are not empirically equivalent to within σ.

It is tacitly agreed, in all factual sciences, that empirical equivalence amounts to empirical confirmation, and inequivalence to disconfirmation. Therefore we adopt the postulate that a hypothesis is *confirmed* (or *supported*) by all the data that are empirically equivalent to it, and *disconfirmed* (or *undermined*) by all those that are empirically inequivalent to it. It is also tacitly agreed that, the more numerous and varied the empirical evidence favoring a hypothesis, the more strongly it is confirmed.

A number of philosophers have attempted to quantitate the concept of degree of confirmation (or corroboration) of hypotheses in terms of their probabilities or, alternatively, improbabilities. In fact an entire industry has developed out of those proposals. But, since the assignment of probabilities to hypotheses is arbitrary, such measures are arbitrary and therefore useless at best, misleading at worst, and in all cases alien to the ideal of objectivity. In science and technology probabilities are assigned only to random *events*, never to propositions. What is true is that probabilities of random events are sometimes used to evaluate hypotheses holding that certain events are not random. In this case the probabilities of events, not of hypotheses, are compared. An explanation of this point may be in order.

Consider the process of self-assembly of a macromolecule, such as a gene, out of its precursors or components. If one models such components as point particles colliding at random, one is able to estimate the probability that the macromolecule got assembled by chance. Since the resulting number is extremely small, one concludes that the precursors did not assemble by chance. If one believes in miracles, one assigns the formation of the molecule to supernatural agencies; otherwise one explains it in terms of interatomic and intermolecular forces such as hydrogen bonds. This kind of reasoning is done all the time in psychology and other sciences, and it has been polished by mathematical statistics. The key words are 'null hypothesis' and 'significance level'. A quick reminder may be welcome.

Suppose we are to test a hypothesis of the form "B depends upon A" e.g. "A causes B". The thing to do is to compare the test results with the rival hypothesis according to which the occurrence of B is a chance event independent of A: this is the *null hypothesis*. *Example*: An animal can press either of five keys, every one of which has a different letter painted on it; the experimenter wishes to test the hypothesis that the animal discriminated among them and has a definite preference for one of them. The null hypothesis is of course that the animal presses the keys at random, so that the probability of each letter being pressed is 0.20: this is the so-called *chance level*. If the animal does press any one letter with a frequency significantly greater than 0.20, i.e. *above the chance level*, the null hypothesis has been refuted and, by the same token, its rival has been confirmed: i.e. the animal has in fact learned. The significance level (p) is a measure of the "significance" of the above-chance result and so of the rigor of the test. More precisely, p is the probability of observing a difference with respect to the chance level when actually there is none. The value of p is adopted before the experiment is run and is usually required to be 0.05. This means roughly that, if the null hypothesis (chance) is true, then the data will exhibit a difference (with respect to the chance level) only once in every 20 times.

One point to note here is that neither p nor any other probability involved in the above test is the probability of a hypothesis or the degree of confirmation of it. Another point is that the experimenter, far from trying to disprove his own hypothesis, *attempts to refute its rival*, i.e. the null hypothesis. In so doing he inverts Popper's injunction. But of course he takes precautions: by adopting a low significance level, such as $p = 0.005$, the experimenter makes sure that it is unlikely that he will adopt his own hypothesis when in fact the null hypothesis is true. So, the test he gives his pet conjecture is severe although he makes no attempt to refute it: he aims for a *rigorous confirmation.*

At first sight a reasonable measure of the degree of confirmation (or disconformation) of a hypothesis by a set of data would seem to the ratio between the difference and the sum of the number of confirmers and that of refuters of the hypothesis, i.e. $c = (C - R)/(C + R)$. The value of this function is 1 for $R = 0$, $\frac{1}{2}$ for $R = C/3$, 0 for $C = R$, $-\frac{1}{2}$ for $R = 3C$, and -1 for $C = 0$. However, since these values obtain regardless of the total number $C + R$ of tests, as well as of their variety and rigor, they are hardly significant. In conclusion, the above formula is useless. The alternative formulas proposed by inductive logicians fare even worse.

Scientists, usually eager to pin numbers on almost anything, are wary of quantitating the degree of confirmation of hypotheses. Thus nobody in his right mind would say, e.g. that the degree of confirmation of the Schrödinger equation is $\frac{3}{4}$. All we find in the scientific literature is a ranking of empirical evidence strength: *very strong* (or *compelling*), *strong* (or *definite*), *weak* (or *presumptive*), *very weak* (or *indecisive*), and *nil* evidence for or against a hypothesis. Thus the hypothesis that terrestrial life emerged about 3,000 million years ago is judged to be very strong; that mental functions are brain processes, strong; that the neutrino has a nonvanishing mass, weak; that intelligence is inheritable, very weak; that there are superluminal particles, nil; and that acquired characters are inherited, very strong negative.

The weight of a body of empirical evidence depends on the status of the

TABLE 12.3
Impact of new empirical evidence on evaluation of hypothesis

Cases	*Intrinsic value*	*Effect on scientific community*
e confirms established *h*	Minute	Nil
e undermines established *h*	Medium	Weak
e confirms new unrivalled *h*	Great	Strong
e undermines new unrivalled *h*	Small	Weak
e confirms new rival *h*	Great	Medium
e disconfirms new rival *h*	Medium	Weak
e inconclusive for or against established or new *h*	Nil	Nil

hypothesis it is relevant to: it is not the same for an established conjecture as for a new one. And the impact of new evidence on the scientific community depends not only on the intrinsic value of that evidence but also on a number of cultural factors. For example, the scientific community does not hail as enthusiastically as the mass media reports of the refutation of the theory of evolution, or of the confirmation of the hypothesis of telepathy. Table 12.3 exhibits the main situations with regard to both intrinsic value and social impact.

As can be seen from this table, the value of an empirical finding depends crucially on its relation to the hypothesis to which it is relevant. If the results are inconclusive one way or another, it is advisable to alter the experimental design so as to enlarge the sample, or render it more representative, or reduce the constant and random errors. The findings are of maximal value if they support a new unrivalled hypothesis or a new rival. They are far less weighty if they confirm or undermine an established hypothesis. In particular it takes more than a few unfavorable data to bring down a previously confirmed hypothesis. One does not rush to proclaim its demise but tries again, taking extra precautions or even altering radically the experimental design. This is particularly so if the hypothesis happens to belong to a theory enjoying empirical supports of several other kinds. In contrast, it takes only one conclusively negative finding to refute, at least for the time being, a new rival to an established hypothesis. (The latter may enjoy such an enormous prestige that any rival to it may be regarded as a crank not worthwhile testing—which is just as well in the majority of cases though unfortunate in a few.) And it takes many positive tests to conclude to the truth of a hypothesis.

How decisive can empirical findings be? In particular, are there crucial experiments? If by 'crucial' is meant a conclusive and therefore definitive verification of a given hypothesis or theory, then the answer is in the negative, for future alternative hypotheses or theories might account for the same, as well as further, empirical results at least as well. But if by 'crucial' is meant a finding allowing one to *reject* a given hypothesis or theory, or to *prefer* one of them to its known rivals, then there *are* crucial experiments (and measurements and observations). Since existence hypotheses are best proved by exhibiting examples, let us proceed to do so.

Uncounted observations and measurements have refuted conclusively the hypothesis that the Earth is flat; by the same token they have confirmed conclusively the hypothesis that it is (roughly) round—so much so that we call this a fact. Likewise uncounted observations have refuted conclusively the hypothesis that the biospecies are fixed; and, *eo ipso*, they have confirmed conclusively the hypothesis that biospecies have evolved—though the exact manner of evolution is still *sub judice*. The cases of the hereditary transmission of acquired characters, of demonic possession, of the biological inferiority of certain human groups, of the immutability of the human mind, and of the permanence of all social institutions, are parallel: they have all been refuted conclusively and, by the same token, their negations have been confirmed conclusively. (However, note that it is far easier to confirm a negative proposition than a positive one. Thus the conjecture that the reader of this page is exactly 30 years old is likely to be false though perhaps not widely off the mark. Most readers will be instances of the statement that any present readers are not exactly 30 years old.) In short, the radical skeptic is wrong: empirical evidence *can* be found to refute hypotheses in a conclusive manner and thus to confirm conclusively their negates. It is not true that all hypotheses are equally wobbly and stay so forever, so that nothing is ever settled. For better or worse we do not have to start from scratch every morning.

As with hypotheses so with hypothetico-deductive systems: there *are* observations (or measurements or experiments) capable of tipping the balance, in a decisive manner, between two rival theories. One case will suffice to prove this existence hypothesis. Hertz's experiments on electromagnetic waves destroyed once and for all the action at a distance theories of electromagnetism existing in his time and showed the enormous superiority of Maxwell's field theory over them. This is as close to a proof as one can get in factual matters: it proved that theory T_2 (Maxwell's) was *truer* than theory T_1 (e.g. the Gauss and Weber theory)—without however proving that T_2 was true. But it so happened that, no sooner was T_2 accepted as *the* true theory of the electromagnetic field, than Planck found a flaw in it (blackbody radiation) and Einstein another (photoelectric effect). This showed that T_2 was in need of basic corrections. These were provided by quantum electrodynamics (T_3), which is not only more accurate but also far richer than T_2, and consistent with quantum mechanics to boot. Still, T_3 is known to have certain flaws (the so-called divergences) that call for the construction of an even truer theory T_4 which is still due. In conclusion, although no experiment (or measurement or observation) can conclusively verify any theory, some experiments allow one to make *decisive choices* between given rival theories.

Why cannot theories (hypothetico-deductive systems) be confirmed in a conclusive manner: why must we settle for choosing the best confirmed member of a pair of theories? One reason is of course that, unlike a hypothesis, a theory is a system of infinitely many statements. (Recall Ch. 9, Section 1.2.) We could not possibly check every one of them but must select, for test purposes, a finite number of hypotheses in the theory. So, even if all of these statements were to pass the empirical tests, there would still remain infinitely many untested propositions in the theory. The second reason for the impossibility of confirming conclusively (i.e. verifying) a theory is that, for it to be checked, it must be enriched with data and indicator hypotheses (Ch. 11, Section 2.2) and, possibly, with subsidiary hypotheses as well. In other words, it is not just the bare theory but the theory together with all such additional propositions—some of which are data and others hypotheses—that is confronted with the new empirical evidence. Consequently we may blame any unfavorable outcome of the test on some of the additives to the theory, and may propose to revise some of these. And if the test is favorable to a particular theory-data-indicator hypotheses-subsidiary hypotheses combination, we must still try further such combinations with the same theory. Should all of them succeed we may hazard to diagnose the theory as (sufficiently) true. However, strictly speaking no theory can be assigned a truth value, for such assignment would require checking its infinitely many formulas.

Popper (1959) has held that confirmation is worthless because it is easy to come by, whereas refutation requires severe tests and can be decisive. A number of cases confirm this view, but there also arguments in favor of the thesis that confirmation and refutation are equally cheap or costly depending on the background knowledge used to design and interpret the tests as well as on the severity of the latter. Let us first offer three cases that seem to lend inductive support to Popper's anti-inductivist methodology.

First case: "My dog keeps the ghosts away by wagging his tail" seems to be confirmed by the nonexistence of ghosts in his vicinity. (Actually there has been no experimental test of the hypothesis, for the dog has not been prevented from wagging his tail, and there has been no independent check on the ghosts hypothesis.) *Second case*: "Of the two hypotheses, creationism and evolutionism, the former is by far the better confirmed, for whatever happens is in the last resort the will of God". (Likewise no controls have been set up. Moreover it would be considered blasphemous to try and control or even monitor in any way the divine thoughts and deeds.) *Third case*: "There is plenty of evidence for witchcraft, such as the hundreds of confessions signed by witches during the 16th and 17th centuries, not only under torture but also spontaneously, and all of which read very much alike—a likeness that cannot be mere coincidence". (Needless to say the Inquisition did not conduct any experiments on the presumed witches, nor were these examined by 20th century psychiatrists to find out whether they were psychotics.) This case is of great methodological interest because it shows the crucial role played by the "interpretation" of data in the light of our background knowledge. Indeed, it appears that the said confessions were cast in the language of the demonology concocted by Dominican priests on the basis of pagan myths and Christian theodicy, much as our contemporaries undergoing psychoanalytic treatment talk, feel and dream as dictated by their therapists. Thus "the mythology created its own evidence, and effective disproof became ever more difficult" (Trevor-Roper, 1969).

The above examples confirm, then, Popper's thesis that empirical confirmation is insufficient for it can be had too easily for any hypothesis. But they also show what is wrong with his thesis that confirmation is unnecessary and that only refutation, or unsuccessful attempts at refuting, count. In the first place, all three cases discussed above were examples of observation, not experiment: if the proper controls had been incorporated into the observations, the results would have been different or, rather, there would have been no favorable results. Second, all empirical operations are dependent upon our background knowledge: data do not come bare but wrapped in presuppositions. By altering some components of this wrapping—some of the hypotheses or data included in it—we may be able to change a confirmation into a refutation or conversely. Thus if we believe in witches we shall tend to accept reports on witchcraft, whereas if we disbelieve in witches we shall discard them out of hand—or shall inquire into the mental health or the level of education of our informant. The point is that unfavorable empirical evidence is just as vulnerable as positive one to the falsities in the background knowledge against which it is evaluated.

There are further arguments against refutationism. One is that it is far from easy to confirm a precise theoretical prediction, such as that the 3489 Å (ultraviolet) line in the spectrum of lithium results from a transition between two bound excited states of the negative lithium ion (Carlos Bunge, 1980). Checking this prediction requires constructing a device for producing negative ions in a certain state and using ultraviolet spectrometry techniques. A second argument is that it is psychologically and methodologically false to pretend that scientists value only attempts to shoot down their favorite hypotheses. To be sure, one will examine critically his theories but, once built and checked, it is only human to wish to see them reconfirmed. Moreover, when convinced of their truth one will tend to disregard criticism. Thus when the noted experimentalist W. Kaufman announced in 1906 that his experiments had refuted the special theory of relativity, Einstein and Planck did not budge. Ten years later it was discovered that the experimental results had been vitiated by a leak in the vacuum system (Holton, 1973). Luckily neither Einstein nor Planck were falsifiabilists. They were only fallibilists.

As for the methodological need for confirmation, it is just as clear: there is no other empirical indicator of factual truth. This is why physicists are still withholding judgment concerning the hypothesis of the existence of gravitational waves, black holes, and magnetic monopoles until someone detects them. No theoretical argument and no criticism can substitute for clear-cut—even if revisable—empirical confirmation.

The above reasons suffice to reject refutationism (deductivism) as well as confirmationism (inductivism). Section 3, on the nonempirical indicators of truth value, will yield further arguments against those two doctrines. The view we adopt instead can be summarized as follows. First, confirmation and refutation usually come in degrees such as nil, very weak, weak, strong, and very strong. Consequently factual truth itself comes in degrees. Second, because hypothetico-deductive systems contain infinitely many propositions, and some of them (namely the data, indicator hypotheses, and subsidiary assumptions) originate outside the systems, these can never be fully tested. However, if two rival theories are testable at all then we can, at least in principle, find out which of them is the best confirmed by a given set of data. Thus a choice among theories is possible. Third, imprecise hypotheses and theories withstand empirical tests far better than precise ones. In particular the negation of an exact hypothesis—such as "$a \neq b$"—is in general easy to confirm and therefore of little use except as a guide in the search for alternatives. Negation is cheaper than affirmation.

Our fourth principle is that failure to refute a hypothesis may not be of much value, particularly in the case of existential hypotheses such as those affirming that there are black holes, vector bosons, and extraterrestrial intelligent beings. Indeed failure to find the thing conjectured may be blamed on unsuitable equipment or search tactics. Fifth, although confirmation is insufficient, it is indispensable for assigning (temporarily) truth to some degree. Sixth, every scientist subjects his conjectures to critical examination and asks his peers for comments and criticism on them, but he also wishes to defend them from destructive criticism if he believes them to contain some grain of truth. In criticizing a theory we should be rather generous, in defending it rather parsimonious: the former because we do not want to nip possible advances in the bud, the latter because we do not want to condone error. Seventh, the value of confirmation and refutation depends critically upon the quality of the antecedent knowledge employed to produce and interpret the data concerned. Eighth, in actual scientific practice hypotheses and theories are evaluated not only in the light of data but also in that of nonempirical criteria, such as consistency. This last point will concern us in Section 3.

2.2. *Efficiency*

Cooking and medical recipes, technical rules and moral norms, as well as instructions and programs, are not propositions but *proposals*. And they are not true or false but efficient or inefficient. (On the other hand a proposition of the form "Rule r is efficient" may be true or false.) However, the adequacy of proposals is (or ought to be) assigned on the strength of tests. Not tests for truth, to be sure, but tests for efficiency. Thus we want to find out whether a given prescription for curing a certain illness is efficient, and whether a certain social program for eradicating a given social plague works.

In principle, any given proposal may be efficient to some degree, or inefficient, or even counterproductive. In some cases we may quantitate such efficiency, i.e. we may assign a proposal an efficiency value comprised between -1 (maximally counterproductive) and 1 (maximally efficient).

And the efficiency of a course of action ruled by the given proposal, in attaining its goal, can in turn be set equal to the product of the efficiency of the proposal by the value (or utility) of the goal itself.

Practice, which is not a test of truth, is the ultimate test of efficiency. Hence efficiency tests must be given independently of truth tests. This is not, of course, the view of pragmatism and dialectical materialism, for which only practice counts. But this view is wrong, as shown by the uncounted theories in basic science that are practically useless and therefore neither efficient nor inefficient, as well as by the uncounted myths that have been highly successful—e.g. in recruiting believers and activists—without being true.

In general, then, efficiency is not equivalent to truth, let alone identical with it, and the two predicates are not related in any regular way. There is, however, one important case where such relation does exist, namely when a proposal is based on scientific laws. For example, if we have ascertained that events of type *A* are always followed by events of type *B*, and estimate that we can control or produce events of type *A*, then we can set up the proposal "In order to get *B*, do *A*", or *B per A* for short. As a matter of fact the same law justifies also trying the dual proposal "In order to avoid *B*, prevent *A* from happening", or *non-B per non-A*. That is, one and the same set of laws may be used as a foundation for a pair of proposals, one for attaining a goal and the other for avoiding it. We call such proposals *nomologically grounded* (or *justified*). And we stipulate that technology fashions and recommends only nomologically grounded rules, norms, and proposals. (More in Bunge (1967b).)

Note that basing a proposal on a set of laws bears on efficiency, not morality: thus the proposal might be the efficient killing of innocent people. Hence showing that a proposal is based on a set of laws justifies it only partially. A full foundation of a proposal to use means *A* to bring about goal *B* must include a moral justification of both means and goal. Likewise if we propose to prevent *A* from happening in order to avoid *B*, we must justify morally our not doing *A* as well as avoiding *B*. More in a while.

The differences and relations between a law and the proposals based on it are summarized in the comparison between the truth table of the former and the efficiency table of the latter: See Table 11.3. In it we have assumed, for the sake of simplicity, that the law statement "If *A* then *B*" is completely true rather than only partially true. (This form fits both causal and probabilistic laws. In the latter case the consequent reads "*B* with probability *p*".) And we have assumed that the efficiency of the correspond-

TABLE 12.4
Truth table of law and efficiency tables of rules based on law.

A	*B*	*If A then B*	*B per A*	*non-B per non-A*
1	1	1	*f*	—
1	0	0	0	—
0	1	1	—	0
0	0	1	—	*f*

ing proposal can be gauged by a number *f*. We have not set *f* equal to unity because a law statement usually refers to an idealized (in particular simplified) model of things of some kind. Such ideal conditions can often be approximated in the laboratory. On the other hand real life, which is the domain of practical proposals, is messy. Hence the efficiency of a proposal may be less than maximal even if the underlying law statement is extremely accurate. This does not entail that practical proposals can dispense with law statements. On the contrary, they have to involve further law statements to account for the complexity of real life situations.

Note the following additional differences between proposals and their underlying law statements exhibited by Table 12.4. To begin with, whereas the law statement "If *A* then *B*" has a truth value for all four combinations of antecedent and consequent, the corresponding proposals have efficiency values only in half of the cases, namely when the means *A* is in fact applied. In this case, if the goal *B* is achieved to the extent *f*, we infer that the proposal works to precisely this extent. But if *A* is not implemented then we cannot infer anything with regard to the efficiency of the rule.

Finally, whereas law statements are value free, proposals are not, since they all concern goals that are more or less valuable. Moreover the means as well as the goals may be the subject of valuation, not only economic but also moral. Therefore mastering a law of the form *If A then B* does not warrant jumping without further ado to the rules ***B per A*** and ***non-B per non-A***. We must also evaluate both *A* and *B*, and we should declare explicitly what our preferences are. Our conduct should then be guided by the following maxim: Try ***B per A*** if, and only if, (*a*) "If *A* then *B*" is a law statement, (*b*) *A* is attainable or feasible, and (*c*) *B* is desirable and moreover outweighs *A* by far. The dual role, *non-B per non-A*, is subject to similar conditions. Condition (*a*) is epistemological, (*b*) is technical, and (*c*) is axiological (possibly moral).

The above maxim starts with the word 'try' because the conjectured efficiency of a rule should be checked rather than being taken for granted. As suggested before, a proposed rule may have to be altered (in particular complicated) to cope with reality. Or it may have to be replaced with an alternative rule, in particular one involving a different means to the same goal. And different rules are likely to have different efficiencies. The truth of a law statement suggests the efficiency of the associated proposals but does not guarantee it nor, in particular, does it guarantee maximal success. Nor does success warrant truth: it only poses the problem of explaining why the given proposal works. Thus the apparent success of acupuncture in decreasing pain poses the scientific problem of uncovering the neural mechanism whereby acupuncture raises the pain threshold.

In conclusion, there is no simple relation between truth the efficiency. Consequently the relations between theoretical knowledge and practical knowledge, science and technology, and knowing and doing, are involved. In particular, technology does not flow automatically from science—nor the other way round. (Obvious moral for science and technology policy: Cultivate both.) Another consequence of the complex relation between truth and success is the inadequacy of pragmatism as a philosophy of either science or technology for, by conflating truth and efficiency, pragmatism accounts for neither.

3. CONCEPTUAL VALUE INDICATORS

3.1. *External Consistency*

Empirical confirmation, though necessary, is not a sufficient indicator of factual truth, and this for the following reasons. Firstly, in principle any number of hypotheses and a large number of theories can cover the same data. I.e. two or more inequivalent hypotheses or theories may enjoy the same inductive support—as was the case in the 16th century, though not in the next, with the sun-centered and the earth-centered planetary astronomies. Secondly, the zeal in fitting the data may be overdone. Thus by increasing the number of uninterpreted (phenomenological) parameters one can improve indefinitely the fit to any given set of data, though at the price of losing understanding and predictive power: overfitting does not pay.

In short, data are insufficient grounds for choosing among competing hypotheses or theories. Nor is internal (logical) consistency sufficient in the case of theories. Though necessary, this condition too is insufficient, for one can spin consistent theories at odds with the facts. This being so, there are two possibilities: either nothing can help us choose between two consistent and equally well-confirmed theories, or something can. To find out which is the case we must examine the way scientists and technologists proceed—rather than, say, consult some philosophical authority. This section and the next will be devoted to such examination.

When an investigator sketches his research plan—in modern parlance when he 'writes his grant application'—he makes sure that, no matter how novel his ideas may be, they do not clash with the whole of antecedent knowledge. To put it positively, our researcher takes it for granted, or attempts to show, than even his most heterodox ideas and procedures are congenial with our background knowledge or, rather, with the bulk of it. For example, if he proposes to investigate the hypothesis that inoperant ("selfish" or "junk") genes are fossil ones, i.e. nonfunctional vestiges of long past evolutionary stages, he will not question current mathematics, physics, chemistry, or even the bulk of genetics, let alone evolutionary biology.

We may say that the scientific investigator strives to keep or attain *systemicity* and, in particular, *external consistency*, or the compatibility of his new views with the bulk of antecedent knowledge (Bunge, 1967b). He may well be a revolutionary in one department but he cannot hope to upset the whole of science at one stroke—unless he is a pseudoscientist. The same holds for technological research, design, and planning, which are increasingly being based on applied science. True, the technologist may occasionally defy and refute the odd scientific theory, as was the case with the first aircraft designers, who had to ignore the primitive aerodynamic theory then prevailing. But, unless he is a bogus technologist, he will not go against the grain of mathematics and science; moreover, he will avail himself of as much applied science as he can lie his hand on. (Not doing so can be ruinous, as Edison learnt when he ignored the experts who had told him that, according to theory, alternating current was more efficient than direct current for transmitting electric energy.) In short, the technologist too values external consistency, and does so increasingly.

Why external consistency? Why could we not question everything at the same time? My answer is as follows. First, we *need* not question the entire body of knowledge: we question items of antecedent knowledge only as we find fault with them. Second, what we do need is to extend and deepen the existing knowledge, i.e. to face new problems. (Sure enough, as we proceed we are bound to revise some of the very body of knowledge that we used as our starting point.) Third, we need a launching pad, however wobbly it may be, for no problem can be posed in a vacuum. Every problem has presuppositions—which, of course, may be questioned provided we take some other propositions for granted. (Recall Ch. 7, Section 4.1.) Fourth, by using—at least temporarily—some of the antecedent knowledge we are assured of a fair amount of heuristic guidance and indirect empirical support. For example, if a new hypothesis implies or is implied by another hypothesis that has already passed empirical tests, then any direct support for each of them is indirect support for the other: see Figure 12.3. Fifth, even if we wanted to revolutionize the entire body of antecedent knowledge we *could* not do it, not only because every problem has presuppositions, but also because the evaluation of new ideas cannot be done in a vacuum and must not be done arbitrarily. Thus we shall be justified in preferring one theory to another only if the former accounts better for data or hypotheses that are not questioned for purposes of evaluation. Every well-grounded preference judgment is of the form "*A* is preferable to *B* with regard to (or in view of, or to attain goal) *C*", where *C* is not questioned during the comparison although it may eventually become an object of criticism. For all five reasons, although science and technology are in permanent transformation they cannot suffer total revolutions. We shall return to this point in Ch. 13, Section 3.2.

Let us watch the principle of external consistency at work. *Example 1*: Wundt (1879) rejected spiritism not for the dearth of empirical data in its favor but because it contradicts the best established scientific laws. (He also noted that magicians and mediums are better equipped than scientists to disclose the tricks of their colleagues, partly because scientists are used to acting in good faith and do not suspect any trickeries.) Hebb (1952) used likewise what he called the "external criteria" of physics and physiology, which "say that ESP is not a fact despite the behavioral evidence that has been reported". *Example 2*: Every time a new theory is proposed in place of an established one, it is judged not only as to internal consistency and empirical support, but also as to its harmony with the bulk of antecedent knowledge. Moreover the new theory is required to yield, in some limit or other, roughly the same correct results as its rival. Physicists call this methodological rule the *correspondence principle*, and most of them refuse even to consider any theories violating this requirement. And for a good reason, namely that external consistency, and in particular "correspondence" with a reasonably good theory, is a weighty truth indicator.

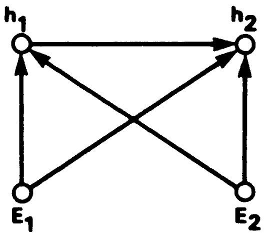


Fig. 12.3. If the hypotheses h_1 and h_2 are logically related, then every confirmer (or refuter) of one of them confirms (or refutes) indirectly the other.

Of course there is risk in applying the criterion of external consistency. By invoking it we may occasionally thwart a valuable piece of heterodoxy—but then error is the unavoidable hazard of inquiry. The principle serves not only to evaluate scientific hypotheses and technological proposals: it also spares us uncounted disasters by weeding out wrong-headed (ill founded) research projects and entire pseudosciences. To give it up would be to condone bogus science and deprive the serious investigator or inventor of any base line to start his inquiry, as well as of a guide to evaluate research plans and outcomes. It is no accident that Feyerabend (1975, Ch. 3) rejects the principle of external consistency and places magic, religion and pseudoscience on the same footing as science.

To prevent the principle of external consistency from enshrining dogma and stifling research it suffices to demand that (*a*) research produce new results—a standard requirement, and (*b*) our fundamental principles be subjected to critical scrutiny and systematization once in a while. The latter task is called *foundations research.* It is an established branch of mathematics but it still has to win acceptance in factual science, where observation and calculation are far more highly esteemed than discussion of basic principles.

The neglect of foundations research is mistaken, for unexamined principles may hide error or may need the company of further principles. And dismissing such research as "mere philosophy" betrays superficiality, for some fundamental principles of science and technology are in fact philosophical. For example, every fundamental physical hypothesis is required to comply with general philosophico-scientific principles such as those of antecedence ("Effects cannot precede their causes") and invariance with respect to observer changes as well as changes in the observer's standpoint or reference system. (For further philosophical principles of science and technology see the Introductions to this volume and to Vol. 3.)

Our view on external consistency as both a necessary requirement and a truth indicator is at variance with three popular philosophies of science: confirmationism (inductivism), refutationism (deductivism), and epistemological anarchism. The first exaggerates the value of examples (confirming instances), the second that of counter-examples (disconfirming instances), the third cares for neither ("Anything goes"), and all three views ignore the matter of external consistency. Yet even the most cursory examination of scientific practice shows that (*a*) data are not compelling by themselves (inductivism) or accepted by convention (deductivism) but only provided they cohere with some previously accepted data and theories; (*b*) a theory is not accepted just when it enjoys a strong inductive support (inductivism) or when all attempts at refuting it have failed (deductivism), but only when these two conditions have been met and, besides, the theory is consonant with some other theories (particularly theories about lower level entities); (*c*) the slogan "Anything goes" may be justified in matters of taste (though even here there are some restrictions) but not in matters of truth (science) or efficiency (technology).

In short, we postulate that every new proposition or proposal be compatible with the bulk (never the whole) of our antecedent knowledge. Moreover we submit that modern science and technology in fact abide by it. This postulate has the following consequences. First, failure to comply with it is a clear indication of pseudoknowledge—in particular bogus science or technology. Second, even though science and technology are characteristically changeable, there can be no total revolutions in them: every transformation, however deep, is local. Nearly total revolutions were possible before the emergence of modern science and technology: from then on every epistemic revolution has been and will be partial because it is a new growth in the midst of a vast system of knowledge. Third, because every body of background knowledge has philosophical (in particular ontological and epistemological) components: (*a*) the evaluation of every radically new proposition or proposal involves or presupposes philosophical considerations; (*b*) there are no frontiers between (good) philosophy on the one hand and the rest of knowledge (in particular scientific and technological) on the other; (*c*) the search for a demarcation criterion between philosophy on the one hand and science and technology on the other is futile (Bunge, 1973b); however, (*d*) since most of philosophy is still at variance with or at least remote from contemporary science and technology, only a deep and sweeping revolution in philosophy or an equally deep and sweeping counter-revolution in science and technology can bring about an overall consistency between philosophy on the one hand science and technology on the other.

3.2. *Other*

Internal and external consistency are not the sole nonempirical indicators of the value of propositions and proposals. Several others, some reliable and necessary, others less reliable and less desirable, or even ambivalent, have been proposed by a few scientists and philosophers. We shall list them with the help of an earlier study (Bunge, 1967b, Ch. 15, Section 15.7). Let us start with the reliable indicators, labeled '*R*'.

In the first place we have three *formal* or logical indicators of the value of scientific or technological hypotheses, theories, or proposals:

R1 Well-formedness: the propositions and proposals should be well-formed rather than gibberish. This condition excludes expressions such as 'The length of *b* is 3′, where the mention of a unit is missing, and 'That method is good', which fails to mention the goal.

R2 Systemicity: it is desirable that every construct be a system (classification or theory) or a component of it. This excludes strays, which are hardly intelligible and cannot be evaluated for truth or efficiency.

R3 Internal consistency: propositions and theories should be logically consistent. This excludes contradictions and demands that we reconstruct any theories containing them.

Next comes a *semantic* indicator:

R4 Interpretability: some mathematical formulas in a scientific or technological theory or proposal should be susceptible to interpretation in factual (though not necessarily empirical) terms. To this end we must join it to semantic assumptions determining what the various predicates refer to and represent. (For semantic assumptions see Ch. 9, Section. 2.2.)

In the third place we have a rather numerous collection of *epistemological* indicators, other than coverage and predictive power, which may be regarded as empirical:

R5 External consistency: i.e. compatibility with the bulk of antecedent knowledge: see Section 3.1.

R6 Consilience: "the evidence in favor of our induction [theory] is of a much higher and more forcible character when it enables us to explain and determine cases of a *kind different* from those which were contemplated in the formation of our hypothesis [...] No accident could give rise to such an extraordinary coincidence" (Whewell, 1847, Bk. XI, Ch. V, Section 11, p. 65).

R7 Unifying power: capacity of a theory to embrace previously separate theories—e.g. electromagnetic theory of light, neurophysiological theory of mind—or of a plan of action to bring together previously detached courses of action—e.g. a social program aiming at both job retraining and public works aimed at alleviating unemployment. Clearly, unifying power entails consilience.

R8 Depth: deep theories and techniques, involving fundamental mechanisms, are preferable to shallow ones. In science because deep theories give more than coverage, namely explanation; in technology because deep proposals expose the roots and allow us to control or alter them.

R9 Stability: the theory should not fail at the first unfavorable evidence, and the plan should not founder at the first practical obstacle, but both should be susceptible to repairs. However, stability has a limit. We do not want to save a theory by evasive manoeuvres such as adding further epicycles, for it may then become impregnable to disproof. Nor do we wish to save a plan by making so many compromises that it will not attain its goal.

R10 Heuristic power: new theories and techniques should suggest and guide research rather than just summarizing it, let alone blocking it.

R11 Originality: bold (yet not wild) theories and methods, enabling one to explore old territories in new ways, or to discover new ones, are preferable to well-tried and safe ideas. After all, the goal of research is to gain new knowledge or new power.

Now come two *pragmatic* indicators, particularly desirable in the case of applied science and technology:

R12 Methodological simplicity: the theory should not be so complicated that any calculations with its help, or any empirical tests of it, would be so time consuming that in practice it would remain fallow.

R13 Feasibility: techniques, designs and plans should be capable of being implemented by real people in real time rather than being suited to ideal people and situations.

Finally come three *philosophical* indicators:

R14 Compatibility with a science-oriented ontology: i.e. one of changing things rather than either unchanging entities or ghostly ones. This disqualifies the hypothesis of telekinesis and the proposal of solving the energy crisis by powering machines by psychic power.

R15 Compatibility with a science-oriented epistemology: i.e. one learning from science and thus capable of stimulating the search for true theories and proposals based on objective knowledge. This condition spares us the attempts to gain truths of fact by purely rational means, or by purely empirical ones.

R16 Compatibility with an ethics: enshrining human rights, in particular the right to know, and human duties, in particular the duty to teach and enlighten.

With two exceptions all of the above conditions are mandatory, hence they are good indicators of truth or usefulness. The exceptions are *R11* (originality) and *R12* (methodological simplicity), which are desirable but not absolutely. Moreover, occasionally they must be dropped altogether. For example, building a theory of the iron atom would not shake the foundations of physics but would still be a very distinguished achievement because of the many subsidiary hypotheses and the enormous computational difficulties involved. On the other hand building a theory of the precise way the nucleic acids synthesize proteins would constitute a revolution and it would probably be a theory difficult to apply as well as to test, just as general relativity was and still is.

A number of alternative nonempirical value indicators have been proposed from time to time, which I submit are unreliable and sometimes clear indicators of disvalue. Here are the most popular, each labeled '*U*'.

U1 Coherence: the articulation of a conjecture with other, previously accepted propositions, is the paramount criterion of truth. This is the thesis of the so-called coherence theory of truth (Blanshard, 1939; Rescher, 1973, 1979). This theory is correct in formal science, where it has been exactified and has become part and parcel of model theory (cf. Vol. 2, Ch. 8, Section 2.1). Moreover it has an important grain of truth with regard to factual statements for, after all, when weighing the truth of a proposition we do so by comparing it with other propositions. (In other words factual truth is always relative to some base line that, in turn, may be questioned as long as a further base line is adopted however provisionally. See Vol. 2, Ch. 8, Section 2.2.) However, internal coherence is insufficient, as shown by a number of consistent doctrines proposed by madmen and pseudoscientists. In the case of factual propositions we need also agreement with facts (correspondence) and external consistency, i.e. agreement with adjoining bodies of knowledge. In short, coherence is *a* truth indicator not *the* truth criterion.

U2 Simplicity: some kind of simplicity, logical, psychological, or practical, has been claimed to be the seal of truth. (*Examples*: the medieval slogan *Simplex sigillum veri*, Mach's *Denkoekonomie*, and the conventionalism of Philip Frank and Nelson Goodman.) To be sure we delight in simplifying, particularly in practical matters. But this is not always possible and it is usually undesirable in cognitive matters because the universe we wish to unveil happens to be complex. This is why the mainstream of scientific and technological progress has been a course of increasing complexity in all regards. Compare contemporary theoretical chemistry, based on quantum mechanics, with Lavoisier's. Besides, high speed computers can handle computational complexities that were formerly intractable: this has brought about a devaluation of simplicity. (For more see Bunge (1963).)

U3 Beauty: theories and methods should appeal to our sense of beauty. Yes, but what is beauty and why should it be related to truth or usefulness? A logically well organized (i.e. axiomatized) mathematical theory is likely to look beautiful to a mathematician or a philosopher, ugly to one interested only in applications, and revolting to a data collector. As with simplicity, beauty is welcome but uninvited.

U4 Longevity: the longer a theory or a technique has "survived" (i.e. been in use), the more adequate or better adapted it must be—so the misplaced evolutionary argument proposed by Peirce goes (see Rescher, 1977). True, some scientific theories (e.g. classical mechanics) and techniques (e.g. microscopy) have proved to be extraordinarily long-lived—but only because they have been refined and assigned more modest roles. If survival were a genuine truth or efficiency indicator, then we should accept as true or efficient all the ancient superstitions that are still with us. No, longevity is not necessarily a mark of truth or of usefulness: it can also be an indicator of ignorance. If anything, the older a cognitive item the sooner it is bound to be altered or even forgotten.

U5 Consensus: the goal of research is to win the consensus of the experts (Ziman, 1968. 1979). This is a double-edged idea. On the one hand it suggests complying with the criterion *R5* of external consistency, but on the other it collides with *R11* (originality) and therefore invites discarding heterodoxies, some of which are likely to carry the seeds of necessary revolutions. Fortunately consensus is not universal in basic science or technology, where controversy flourishes. Even the most widely accepted data, hypotheses, methods, designs and plans meet with criticism. (On the other hand applied science and textbooks are usually free from controversy.) Surely the dissenters are always in a minority: so are the innovators who are eventually proved right.

So much for the nonempirical indicators, both the reliable and the unreliable ones, of truth or usefulness. How are research projects and plans of action *actually* evaluated? What criteria does a supervisor, referee or client actually employ to encourage or discourage a given line of research or development? And how do scientists and technologists evaluate their own findings and those of their peers? There is no consensus on this matter. True, internal and external consistency, originality, testability, predictability (in particular serendipity, or the ability to forecast unexpected facts), feasibility, profitability, and a few other indicators are normally taken into account. In short, evaluation is done on the strength of a whole battery of tests, some empirical, others conceptual. (Only philosophers tend to believe that a single factor, such as simplicity, explanatory power, or predictability, is involved.) However, the entire battery is rarely if ever used explicitly: some indicators are not used, and none is assigned a definite rank, let alone weight. As Kuhn (1978, p. 331) says, they "function not as rules, which determine choice, but as values, which influence it".

Worse, in many instances extracognitive factors play a role in evaluation. Among such factors we may mention personal sympathy or animosity, school affinity or rivalry, excessive depth or originality (incomprehensible to the judge), and even ideological bias. All but the last are easy to understand though hard to forgive. The last factor, namely ideological bias, was highlighted by the Lysenko affair: as is well known the quack agronomist succeeded in beheading Soviet biology by accusing it of not complying with Marxism. Less well known is the case of classical economics, in particular Ricardo's: though at variance with many facts and lacking in predictive power, it dominated the British academic and business communities for an entire century because, among other reasons, it explained social injustice and discouraged social reform (Keynes, 1936, pp. 32ff.). Nowadays neoclassical economics, in particular monetarism, though equally at variance with reality and therefore powerless to predict anything, is upheld because it abhors state intervention and planning. In short, ideological factors do play a role in the evaluation of scientific and technological propositions and proposals. We shall return to this point in Ch. 14, Section 4.2.

The intervention of noncognitive criteria, as well as the intervention of false philosophies of science and technology, may explain the frequent failure of the peer review method. (The method consists in asking experts, whose names are not disclosed, to evaluate the projects and accomplishments of their peers in the same field.) The peer review method is double-edged. On the one hand it screens out much fraud and low grade stuff, particularly most of pseudoscience and pseudotechnology. But on the other hand it tends to perpetuate whatever views the referees happen to hold; in addition, anonymity gives them even the chance to exact revenge. That the method does not work to perfection has been shown by two experiments, which have shown that (*a*) the prestige of an author and his institution are decisive factors in the evaluation (Peters and Ceci, 1980), and (*b*) the fate of a particular grant application is half determined by apparently random elements (Cole *et al.*, 1981). Double-blind refereeing solves some of these problems but not all. In any case no better method of evaluation has been proposed, so we should try and improve on it. One way of achieving this goal is to educate scientists and technologists in the principles of the methodology and philosophy of science, so that they may abandon some of the grossest errors underlying valuational mistakes—such as that data are all that really matter, so that all measurements are worth doing, and no theories are worth considering unless they summarize data or help get more data.

4. CONCLUDING REMARKS

A proposition may be accepted (believed) only if, when subjected to suitable tests, it proves to be sufficiently true. And a proposal ought to be adopted (implemented) only if, on the basis of sufficiently true propositions, it promises to be useful to solve certain problems to a reasonable and good end. But truth and usefulness are seldom manifest: therefore we must resort to indicators.

There are many value indicators, some conceptual and others empirical. (The former may be assessed before any empirical tests are given: such preliminary screening spares much time and effort.) The value indicators may be grouped as follows:

Logical: well-formedness, systemicity, and internal consistency.

Semantical: interpretability in factual terms.

Epistemological: external consistency, coverage, accuracy, consilience, unifying power, serendipity, depth, stability, originality, and heuristic power.

Methodological: testability—i.e. confirmability and, when applicable, refutability as well.

Pragmatic: feasibility (in the case of techniques, plans and designs), methodological simplicity (i.e. simplicity of application), efficiency, and profitability.

Philosophical: compatibility with an ontology of changing things, a realistic epistemology, and an ethics of human rights, in particular the right to inquiry.

Such multiplicity of value indicators renders evaluation delicate but makes for responsible evaluations. In any case that multiplicity ruins the simplistic philosophies of science and technology, such as those summarized in the slogans "*A* theory is only as true as its predictions (regardless of its depth, accuracy, etc.)", and "*A* plan is only as good as its deliveries (regardless of the morality of both means and ends)". The actual process of evaluating propositions and proposals in science and technology is far more complicated than suggested by such simplistic philosophies, but it seldom takes advantage of the full battery of tests, to concentrate only on such properties as accuracy, originality, and usefulness. No wonder, for the scientists and technologists who evaluate the work of their peers have seldom given any thought to the principles of evaluation—principles that they may dismiss contemptuously as being "merely" philosophical.

Unfortunately philosophers retaliate by not doing much better when it comes to evaluating their own ideas. In fact the current standards of excellence in philosophy seem to boil down to smart argumentation (even if it involves ignorance of basic facts), triviality of the problem, possibility of compression into a simple slogan, and prestige of the philosopher's institution. I submit that there is a better way to evaluate a philosophical view, namely to subject it to the following three batches of tests:

Historical: Does the view have any sound historical roots, i.e. does it make contact with the best philosophical tradition? And is it original?

Logical: Does the view constitute a system, and if so is the system internally consistent?

Epistemological: Is the view compatible with the current mathematical, scientific and technological knowledge? Does it solve any interesting problems? Does it have any heuristic power?

When subjected to the above tests, the most popular philosophical views of the day turn out not to be systems, not to harmonize with science and technology, not to possess any heuristic power, and often not even to handle (let alone solve satisfactorily) any outstanding problems. This is a very sad state of affairs which calls for a revolution in philosophy. But the matter of revolutions in knowledge deserves another chapter.

PART V

VARIETY AND UNITY

CHAPTER 13

EPISTEMIC CHANGE

Knowledge, like food, can be stored—for a while. Some of it becomes stale and is eventually recognized as useless, or is no longer of interest to anyone. And whatever does stay is bound to become incorporated into more comprehensive and deeper bodies of knowledge. (Even the immortal Pythagorean theorem has become a particular case of a more general theorem in Riemannian geometry.) Therefore the investigator must be always on the run if he wishes to stay in the same place.

Knowledge is gained and lost in a number of ways: empirically and theoretically, gradually and through revolutions, by disclosing ever finer details and by joining entire fields of research. The more we know the more and harder problems are we able to pose and solve. And the less helpful the existing body of knowledge proves to be to solve new problems, the more it invites its enrichment or replacement.

Like everything new, new knowledge emerges from pre-existing items. And like any other novelty, new knowledge is elicited from within and in favorable external circumstances, and it must overcome resistance. The inner driving force of knowledge is curiosity and, particularly, disciplined curiosity. The external driving force is need and, in particular, the needs of industry, trade, and welfare. And the obstacles to research, hence to novelty in knowledge, are multiple: economic (dearth of resources), political (hostility or indifference to learning on the part of government officials), and cultural (inertia and ideology).

Given the historicity of knowledge, the epistemologist would be ill advised to ignore the history of science, technology, the humanities, and their respective bogus counterparts. And given the social conditioning of research, as well as its impact on modern society, he would be equally ill advised to ignore the sociology of knowledge. However, we cannot find out how knowledge evolves along society unless we know what knowledge is. Therefore epistemology cannot be replaced with the history and sociology of knowledge although it must learn from them.

1. COGNITIVE NOVELTY

1.1. *Cognitive Kinematics*

Recall a few landmarks in the history of human knowledge since the beginning of civilization a few millennia ago: the invention of writing and its use as a thinking tool as well as to record social and natural facts; formal education; the invention of time keeping and astronomy, of mathematics and logic; the substitution of history for myth; the secularization of at least part of the intellectual culture and the accompanying tolerance for some innovation and some dissent; the establishment of scientific and technological bodies, such as academies and observatories; the invention of scientific theory and controlled experiment; the emergence of hundreds of scientific and technological disciplines; and the practical applications of science, i.e. the conversion of arts and crafts into technologies. No doubt, we have come a long way in a period that is but a small fraction of the total time since the human species emerged.

However, by far the greatest number of mathematical, scientific and technological knowledge we possess today has been acquired since the beginning of the Modern Era, which began only five centuries ago. We are all familiar with the exponential growth of mathematical, scientific and technological knowledge over this period. Indeed specialized knowledge has been doubling in size every decade or so, so that its growth may be likened to that of capital at a compound interest of about 7 per cent per annum. The analogy is apt in so far as in both cases the total size at a given moment depends on both the rate of growth and the total size at an earlier time. This highlights the importance of background knowledge, i.e. of the general assumptions and the fund of knowledge. Moreover, we also have the cognitive analogs of bankruptcies and crashes, namely those caused by depressions, wars and invasions, by mistaken science and technology policies, and by irrationalist movements. The exponential growth of knowledge, like that of the world population, is a modern trend, not a law. (Knowledge grows only provided there are people willing and competent to work on new problems and to try new ideas and methods.) Like any other trend, the exponential growth of knowledge may be altered. In fact there are already signs of decline: see Section 4.2.

Mutability is an essential mark of mathematics, science and technology, just as stasis is one of ideology and pseudoscience. This holds for all branches of science and technology, even the classical ones. A look at the

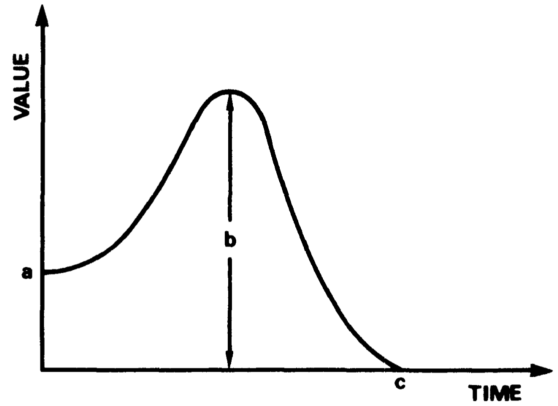


Fig. 13.1 Change in value of a cognitive innovation. Initial value = a; maximal value = b; life span = c.

scientific literature shows that new results keep cropping up even in classical mechanics and thermodynamics, as well as in plant morphology and genetics, in learning theory and human history. It is not only that new data and new hypotheses are found, or new applications are made: some of the very principles of such classical disciplines are modified once in a while. In the case of the classical physical, chemical and biological theories, such renewal is often associated with a modernization of the mathematical formalism. This change is so profound that Newton and Euler would be unable to read a contemporary advanced textbook in classical mechanics.

The value of data, hypotheses and techniques is time dependent. It may increase for a while after inception, reach a peak from which it may decline because the item has been superseded by more accurate information, more comprehensive or deeper theory, or more powerful method—or simply because it has ceased to attract interest. See Figure 13.1. Such decline in value is due to the very nature of inquiry, which consists in investigating problems with the aim of learning something new, i.e. of making original discoveries or inventions, some of which are bound to correct or even eliminate bits of knowledge gained earlier. Obsolescence, not perennity, is the mark of scientific findings and technological artifacts—just as discovery and invention is that of original scientific and technological research and development.

Originality, not only ability to learn much and fast, is the mark of the investigator, be it in science, technology, or the humanities. Originality, feared and repressed until a few centuries ago, is nowadays institutionalized, at least in mathematics, science, and technology. This institutionalization takes the forms of recognition of authorship or patent, and

TABLE 13.1
Phases of normal maturation of science

Phase	*Theoretical aspect*	*Empirical aspect*
0 *Prescience*	Unchecked speculation	Unchecked data
α *Protoscience*	Hypotheses formulated verbally, no theory	Observation and occasional measurement, no experiment
β *Deuteroscience*	Hypotheses formulated mathematically, no theory	Systematic measurement and experiment on perceptible traits of perceptible things
γ *Tritoscience*	Mathematical models (specific theories)	Systematic measurements and experiments on perceptible and imperceptible traits of perceptible and imperceptible things
δ *Tetartoscience*	Mathematical models and comprehensive theories	Precise and systematic measurements and experiments on perceptible and imperceptible traits of perceptible and imperceptible things

academic or pecuniary reward (cf. Merton, 1973). Only in the humanities are endless repetitions of ideas and scholastic commentaries on them, instead of the exploration of new ideas, still tolerated.

We may distinguish five main stages in the normal development of any field of factual inquiry: see Table 13.1. Note the qualifier 'normal': not all fields conform to our schema. For example, cosmology and economics, though mathematically very advanced and well stocked with precise (though insufficient) data, are still dominated by untested speculation; so, they exhibit some signs of maturity together with others of immaturity. (More in Vol. 6.) Also, some fields that teem with unbridled speculation and unchecked data, and so belong in phase zero, are sometimes advertised as scientific although they are clear examples of pseudoscience. (More in Ch. 14, Section 14.1.)

Different fields of research advance at different paces. Thus since mid century biology and history have made great strides, physics and sociology have advanced a fair stretch, psychology and economics only a little. Can the volume and speed of knowledge be quantitated and measured? Several measures have been proposed, such as the number of papers published and quoted. There is even an entire new discipline, scientometrics, with a journal of its own, that purports to gauge scientific and technological advance. However, counting papers and citations gives only an inkling of popularity, not of value.

There is no reliable measure, and no accurate measurement, of anything unless there is some underlying theory. So far we have not got a theory of the growth of science and technology. We only have a handful of rather vague principles, such as the following:

(i) The pace of inquiry in any given field is proportional to (*a*) the fund of knowledge accumulated in the same field, (*b*) the number of fields from which it draws background information, (*c*) the number of workers engaged in research, and (*d*) the number of problems it tackles.

(ii) Research in every field of inquiry can be likened to a river: (*a*) the *mainstream* formed by new lines of research and their continuation, (*b*) the *sidestream* constituted by routine investigations—or, to change the metaphor, it is made up of efforts to fill holes in a terrain the main features of which are already known, and (*c*) the *eddies* composed by barren controversies and heterodoxies (i.e. novel ideas clashing with the bulk of the fund of knowledge).

(iii) The power of any cognitive innovation (datum, hypothesis, method, or design) is proportional to the number of old cognitive items it displaces, and to the number of new cognitive items it renders possible.

(iv) The resistance to cognitive innovation is proportional to the degree of success of the extant body of knowledge which it seeks to displace, as well as to the originality and power of the new item. If there is no rival there is no resistance. If there is, the first reaction to the proposed innovation is to ignore it; the second, to try and find fault with it; the third, to attempt to assimilate it, i.e. to account for it in traditional terms; the fourth, to admit it; the fifth, to turn it into dogma. Resistance to innovation is maximal in stagnant fields. Thus economic theories are often accepted and repeated uncritically, and the economic "establishment" tolerates innovation "of the additive and amending type" but rejects radical novelty (Myrdal, 1973).

(v) Technical innovations are the most readily accepted, the better they satisfy deeply felt needs and wants, and provided they do not harm any powerful vested interests or call for the renunciation of whole life styles.

1.2. *Conceptual Change*

Understandably, philosophers are more interested in theoretical change than in any other type of change: they are interested in the emergence, development, and decline of theories, as well as in their replacement by rival theories. And they are interested in all the aspects of such changes—logical, semantical, epistemological, and pragmatic.

Theories are variously described as ideal objects, systems of changeable meaning and truth value, growing bodies of knowledge, or prescriptions for doing things. These characterizations, though very different, are mutually compatible. In fact from a logical point of view a theory is an ideal object of a certain kind, namely an infinite set of propositions closed under deduction. That one and the same theory may be variously interpreted, and may be attributed different truth values on the strength of different bodies of evidence, is obvious. From an epistemological point of view theories are not static but they grow in certain directions (as more theorems are proved and more applications discovered) and shrink in others (as regions of falsity appear). Finally, from a pragmatic viewpoint theories, even mathematical ones, can be used as rules or prescriptions for computing or for designing experiments, much as musical scores are used as instructions for performing. Since theories are all four—logical objects, semantic systems, growing bodies of knowledge, and prescriptions—according to the viewpoint that is chosen, there is no incompatibility among the four descriptions.

From a logical point of view theories are "already there in one piece" From this vantage point there is no difference between an axiom and the process of hypothesizing it, a theorem and the process of proving it, or a definition and the process of introducing it. Not so from a psychological or a historical point of view. Here definitions are not static identities but either inventions or discoveries of previously hidden identities. Here theorems are not just consequences of a set of premises: they are usually first guessed, then proved. And an axiom is not just an initial assumption that had always been "there" (in the theory): it is a hypothesis, usually tried long after a number of theorems are known. From a logical point of view every result obtained when working on a theory is a discovery not an invention: we only discover occult implication relations. From an epistemological viewpoint there is both discovery and invention: one makes new posits, introduces new concepts, and tries new strategies for constructing proofs.

From an epistemological point of view, a theory is not a self-existing ideal object but a potential body of knowledge that can be partially actualized by working on it. At any given moment we know only a finite subset of all the statements of a theory. This finite and variable collection of statements can be defined as follows. Call T a theory and s a knowing subject. The *part of the theory T known to s at time t* is the subset of the total set of formulas of T that s happens to know at t. And the part of the theory *known to mankind* at time t is of course the union of all such fragments for all the subjects familiar with T. In obvious symbols, $\Theta(s, t) = \{x \in T \mid s \text{ knows } x \text{ at time } t\}$, and $K(t) = \bigcup_{s \in \mathrm{H}} \Theta(s,t)$. Perhaps this, the known part of a theory, is what McMullin (1976) means by a 'dynamic theory'.

A second concept of philosophical interest is that of the family of specific theories built on a given general theory—or, to use the term introduced in Ch. 9, Section 1.2, the family of bound models of the theory. (*Examples*: every quantum-theoretical model of an atom or a molecule is a bound model; likewise every model of the evolution of a biospecies, or rather a biopopulation, is a bound model of the theory of evolution.) Call T a general theory and S_i a set of specific hypotheses and data consistent with T and couched in the language of T. Then the set of consequences of the union of T and S_i is the ith T-model, or $M_i(T)$. And the set of all T-models will be called the *T-family*, or $F(T) = \{M_i(T) | 1 \leq i \leq n\}$. Perhaps this is what Sneed (1979) has in mind when he states (wrongly) that a theory is composed by its hard core and the set of its applications.

Finally consider a historical sequence of theories T_k, each proposed at the corresponding time t_k over a period τ, and each dealing with a given domain (or reference class or universe of discourse) D_k. Moreover assume that the intersection $D = \cap D_k$ of these domains is nonempty. Then the *historical sequence of theories about D* over the period τ is $H = \langle T_k | t_k \in \tau \ \& \ k \in \mathbb{N} \ \& \cap D_k = D \neq \varnothing \rangle$. This seems to elucidate Lakatos's (1970) hazy notion of a research programme. But of course in factual science and technology theories are only part of the evolution of inquiry: there are also empirical investigations.

Broad and deep theories, such as classical mechanics and the theory of evolution, are likely to trigger a landslide of research based on them or inspired by them. Such theories are applied to uncounted new situations and they become paragons to be imitated—in short, they are regarded as paradigms or exemplars (Kuhn, 1962, 1970, 1974). Criticism, whether theoretical or empirical, is unlikely to stop the multiplication of models bound to such generic theories: victorious theories, like victorious politicians, are impatient with critics. Thus hardly anybody listens today to the correct charges of inconsistency made against quantum electrodynamics, or to the objection that "big bang" cosmology rests on flimsy empirical data. The paradigm becomes a paradogma. And, if criticism continues to be unheeded, the paradogma turns into dogma.

Eventually every theory is shown to be defective in some respect or other: in its mathematics, interpretation, accuracy, coverage, or all four. People formerly blinded by the successes of the theory start to admit that it is not perfect, and alterations are tried. Such changes range from minor adjustments to revolutions. In some cases the mathematical formalism is corrected or even thoroughly overhauled, as was recently the case with classical mechanics and classical electrodynamics (Truesdell and Toupin, 1960). In other cases the formalism is kept but the interpretation is altered, i.e. the semantic hypotheses are changed, as is the case with the objective reinterpretation of quantum mechanics (Bunge, 1967c). In still other cases the theory is replaced with another in the same style, as was the case with the change from classical to relativistic mechanics. Finally, in a few cases a whole new style of theorizing is proposed—as when Faraday and Maxwell invented field physics, Darwin and Wallace proposed their theory of evolution, and Lashley and Hebb revived physiological psychology. In such cases learning a new theory is not just adding knowledge but reorganizing from scratch what fund of knowledge one had and adopting a new mode of thinking.

From a semantical point of view theories can be compared as to meaning and truth. According to our semantics (Vol. 1 and 2) meaning is composed of sense and reference. The former is determined by the totality of logical relatives (implicants and implicates) of the construct of interest, whereas the reference of the latter is the class of (actual and potential) entities to which it applies (truly or not). The difference in sense or in reference between two constructs can be taken to equal their symmetric difference. Hence the total *difference in meaning* may be defined as the ordered pair: ⟨Difference in sense, Difference in reference⟩ (Vol. 2, Ch. 7, Section 3.3). In particular, differences in reference can be ascertained rather easily, particularly if the theory has been axiomatized. For example, field electrodynamics refers not only to the currents dealt with by action at a distance electrodynamics, but also to the fields accompanying such currents.

Another respect in which theories are comparable, and are actually compared, is truth. However, such comparison is far more difficult than that of meaning, for it cannot be performed a priori: in fact it involves the calculation of numerous examples and the corresponding empirical tests. Either operation is likely to be more difficult in the case of a rich theory than in that of a poor one: the number of variables and constants is likely to be greater, their interrelations more complex, and they are bound to be less readily accessible to measurement. Thus for several decades Einstein's theory of gravitation was judged superior to Newton's on the strength of only three "effects" in excess of those predictable by the latter. It took several decades of patient calculation of new solutions of Einstein's field equations, as well as ingenious new observations and experiments, to confirm that first impression. Still, the Newton–Poisson theory is so much simpler to apply, and it is such a good approximation, that it is used in the vast majority of astronomical and geophysical calculations. (This fact may come as a shock to the philosophers who believe that Einstein's theory "overthrew" its predecessor.)

In conclusion, rival theories are compared as to meaning and truth; they are also compared as to computational simplicity. Such comparisons are used as a basis for a rational decision among them. Hence it is mistaken to assert that rival theories built in entirely different styles (or involving "Gestalt switches") are mutually "incommensurable", so that the choice among them cannot be a completely rational act (Kuhn, 1962; Feyerabend, 1975). After all, two theories could not possibly compete with one another unless they yielded different answers to equivalent problems—such as alternative explanations of the existing variety of biospecies. And in order to share some problems the two rival theories must share some concepts and, in particular, they must have some common referents.

2. CHANGE MECHANISMS

2.1. *Usual Mechanisms*

Russell asked at one point: How comes it that we are able to know so much even though our individual experience is so limited? Chomsky (1975, p. 7) replied in a Platonic vein: "we can know so much because in a sense we already knew it, though the data of sense were necessary to evoke and elicit this knowledge". That is, experience would be nothing but the Socratic midwife that helps the innate knowledge come to light. There is of course no empirical evidence for this hypothesis. And there is plenty of negative evidence from the neurophysiology and psychology of development: it all points to the thesis that we are born ignorant.

The correct answer is that we are able to know so much because in principle there is no limit to what we can learn by sensing, thinking, and doing. And we can learn so much because our brains are plastic and because we inherit a tradition and learn from our contemporaries. As Bernard of Chartres said, we can see farther than the ancients because we are like dwarfs mounted on the shoulders of giants. That is, we do not have to start our inquiries from scratch because we can make use of the knowledge acquired by millions of people over thousands of years—and because we need not learn it all before tackling new problems. We are born ignorant though with the ability and drive to learn, and we are not born into a social vacuum but into a society that keeps some learning tradition or other. If this tradition happens to be one of inquiry rather than one of reverence for myth, we may learn as much as our genetic endowment and social status will permit.

Next to the innate knowledge thesis comes the combinatory view, according to which every new idea is the result of combining old ideas. There is plenty of evidence for this doctrine, which explains why the more we know the more we can find out. Indeed the more items are available for combination the greater the number and variety of combinations. (Thus if the number of items doubles from 5 to 10, the total number of binary combinations rises from 10 to 45; and if the former increases from 10 to 20, the possible binary compounds rise from 45 to 190.) But of course the combinatory view does not explain the origin of the first ideas or the emergence of radically new ideas. How did the Greeks hit on the idea of a proof: what elements did they combine in order to form it? How did Marx come to believe that human history is nothing but the history of class wars: did he merely mix up some ideas of Hegel's and Feuerbach's? The combinatory view holds only a grain of truth. In particular it applies to the original combination of ready-made parts to build a machine, and to the merger of theories—although it does not explain why some people conceive of such combinations whereas most don't. But combinatorics fails to account for the emergence of radically new ideas, such as those of atom, chemical reaction, gene, and social structure, which do not originate in the senses.

How then is the growth of knowledge, when it happens, to be explained? There are four traditional and influential views on the sources or movers of scientific, technological, and humanistic innovation: empiricism, rationalism, psychologism, and sociologism. According to *empiricism*, knowledge grows spontaneously as experience accumulates either by observation or by chance discovery. Theories would come, if at all, at the end of empirical inquiry and only as data summaries or at most to suggest further observations. Since empiricism regards data as irrefutable and as equally valuable, it imagines the growth of knowledge to be cumulative: it would proceed by the steady accumulation of data. No doubt, the production of empirical data is a necessary component of all factual inquiry; but the invention of theories and methods is another. Nor is it reasonable to doubt that there is some accumulation of data; but many data are in need of revision whereas others, even if accurate, cease to be of interest. In short, empiricism gives a one-sided description of the growth of knowledge: it fits at most the empirical side of protoscience and prototechnology.

On the other hand *rationalism* describes the theoretical side of the most advanced stages of science and technology. According to it, knowledge is self-constructed: it grows by inventing and criticizing ideas. Data would be only occasions for, and checks on, theorizing. The main movers of the inquiry process would be internal: the impulses to question, generalize, systematize, and explain. The growth of knowledge would be a sequence of ever better theories. Once in a while a bold new theory appears that kills its old rival. Knowledge advances thus by fits spaced by uninteresting routine activities. No doubt, this view holds an important grain of truth, but it fails to account for the entire process of research. For one thing fact finding is just as important as theorizing and it engages the vast majority of researchers. For another, new methods, particularly if allied to new and powerful instruments, can be just as revolutionary as new theories: think of the galvanometer, the mass spectrograph, the electron microscope, and the microelectrode. Thirdly, not all of the new and better theories kill their older rivals; thus relativistic quantum mechanics has not displaced the nonrelativistic approximation. In sum, rationalism has nothing to say about laboratory and field research, and gives a distorted view of theoretical progress.

Whereas empiricism and rationalism are popular among philosophers, psychologism and sociologism are often embraced by historians of knowledge. According to *psychologism*, knowledge is always the result of inquiry sparked off by either curiosity or need. Consequently the growth of knowledge must be explained exclusively in terms of mental or biological mechanisms. It would be foolish to deny that psychologism does tell part of the story. But obviously it does not explain why inquiry proceeds swiftly in certain periods and morosely in others, or why it flourishes in certain societies and is repressed in others. *Sociologism* makes up for these shortcomings of psychologism. In its moderate version it reminds us that every inquiry proceeds within a social matrix, so that, far from being a purely individual affair, it engages whole communities of inquirers. In its extreme version sociologism states that every bit of knowledge is a reflection of, or a response to, social conditions: that science, technology and the humanities are the products and mirrors of society. We have no quarrel with moderate sociologism even though it focuses on only one of the aspects of the inquiry process. On the other hand we find extreme sociologism unacceptable: after all, it is individuals, not communities, that perceive, think and act. There is no such thing as a collective brain or mind, just as there is no such thing as a Robinson Crusoe scholar. And a thousand mediocre researchers cannot do the work of a single genius—just as the genius would have nowhere to start from were it not for the dedication of uncounted modest researchers.

The most cursory look at the history of any research line in science or technology shows that each of the four views holds a grain of truth but is otherwise either incomplete or somewhat mistaken. There is no unique prime mover of inquiry: in some cases progress is made by performing new observations, measurements, or experiments; at other times by inventing new methods or new designs; at still other times by inventing new concepts, hypotheses or theories; occasionally also by criticizing or reorganizing known ideas; and always by contrasting ideas with facts, plans and artifacts with values, and so on.

Since each of the above partial views on the growth of knowledge is adequate in some respect, they should be consolidated into a fifth view free from the errors or exaggerations of the other four. This view acknowledges the roles of both experience and reason, as well as of the individual brain and its social matrix. This view results from the *biopsychosociological* approach to inquiry favored in the present book, and it suggests an integrated vision of the history of knowledge. Thus an adequate account of Harvey's work on the cardiovascular system must include such items as Harvey's medical experience and his familiarity with Galen and Vesalius; his acquaintance with the incipient mechanical world view, as well as with some mechanical contraptions, in the first place the pump; his ambition to understand the *humani corporis fabrica* described by Vesalius; his inquisitiveness and audacity (which cost him the loss of many patients); and his living in an age of discovery and invention in a society that tolerated both and was on the way to rewarding them. In other cases the intervention of formal organizations, such as universities, academies, churches, and political parties, must not be forgotten.

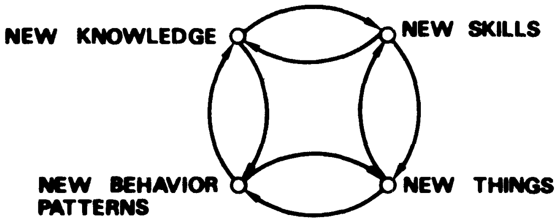


Fig. 13.2. Connections between epistemic and social innovations.

New knowledge sometimes causes social change, and at other times social innovation makes the acquisition of further knowledge possible. Moreover both epistemic and social innovation may result in increasing plant production, which ensues in greater food supply, which makes increase in population possible (not necessary), which causes the growth of cities, which stimulates the enrichment of culture (in particular the invention of new crafts), which favors the evolution of manufacture, trade and services, which in turn offer new opportunities for agricultural development. Likewise, from the 17th century on scientific innovations have sometimes triggered technical changes—such as electrical engineering and industrial chemistry—which have in turn revolutionized production, which has caused cultural and political changes, which in turn have stimulated advances in knowledge. See Figure 13.2. (For the chain-like character of technical and social innovation see Bourke (1978).)

Are such chains necessary? In particular, are discoveries and inventions inevitable? And as a consequence do science and technology have a momentum of their own, i.e. are they unstoppable? All we can say is that every epistemic accomplishment makes further gains in knowledge *possible*. Thus quantum mechanics made it possible to build modern solid state physics, which in turn made it possible—together with electronics—to design high speed computers, which in turn have revolutionized communications, industry and commerce. Without favorable social conditions none of these developments would have happened. They were epistemically possible and socially desirable: this is all we can say.

Yet, it is often stated that simultaneous discoveries and inventions by independent individuals *prove* that they are *necessary*: they seem to be "in the air" and emerge "when the time is ripe". They look as necessary outcomes of earlier work with no need for either genius or chance, or even promise of social reward. So much so that Merton (1973, Ch. 17) holds that what needs explaining is not independent multiple discoveries, for they are the rule, but singletons, which on closer examination should prove to be multiples, as is the case with the rediscoveries of unpublished or poorly circulated work. Merton exhibits an impressive array of data in support of his hypothesis, but does not discuss counterexamples such as Archimedes' statics, Netwon's dynamics, the electrodynamics of Faraday and Maxwell, or Einstein's general relativity. The latter deserves a special comment.

According to his coworker Leopold Infeld, Einstein believed that the special theory of relativity could have been invented by someone else, for a number of people—among them no less than Lorentz and Poincaré—were on the right trail. Not so the general theory. We should believe him, because nobody was interested in gravitation theory at the time: it seemed to pose no problems. Gravitation theory became a hot spot only thanks to Einstein. Even now the observational and experimental "effects" that only Einstein's theory explains and predicts are very small for the most part. Had it not been for Einstein's theory they might have passed unnoticed or, if noticed, they might have been explained in terms of perturbations, experimental errors, or some other *ad hoc* hypotheses.

We should not underrate the roles of genius and of chance. True, an isolated and misunderstood genius will work in vain; still, no scientific community will make great strides unless it contains some exceptionally talented and nonconformist leaders. True, discovery and invention do not spring out of nothing but require some background knowledge; but the researcher may come across some of it by chance—e.g. browsing in the library or meeting a stranger on a plane. In short, new knowledge comes about in ways that are biologically and logically necessary but socially accidental. (See Taton (1955) for the role of chance in scientific discovery.)

What holds for scientific innovation holds also for technical innovation: here too necessity combines with chance, and individual talent and pluck with social opportunity. The importance of the latter is clear from the fact that no industrial or social innovation is the automatic outcome of invention. For invention to fructify in innovation it must be practical and must be understood as useful by those who have the economic or political means to have the blueprints translated into artifacts or social programs. This is why only very few inventions are ever put to use.

To conclude this section: there are as many mechanisms of epistemic change as there are types of epistemic operation. Some workers go out into the field whereas others stay at the laboratory; some make observations and others experiment; some classify and others calculate; some draw blueprints whereas others build theories; some solve problems with the help of existing theories, and others criticize the latter; some delight in specificity, others in pattern; some split and others lump—and so on. The growth of knowledge requires that all of these and more epistemic operations be carried out concurrently in the scientific and technological communities. Breakthroughs may result from any of them. But none will occur unless society tolerates the free search for truth. We shall come back to this theme in Section 4.2.

2.2. *Reduction and Synthesis*

The rise of modern science was accompanied by the substitution of a mechanistic world view for the prevalent organicist one. The new model was the clockwork, not the organism; the new paragon was mechanics, not biology; the new judge was experiment, not dogma. The program of reducing every process to motion triumphed almost everywhere, though not without some philosophical opposition such as Leibniz's. It was not until 1850 that mechanism showed signs of exhaustion, particularly with regard to field physics, evolutionary biology, and social science. Mechanism is now dead. Nowadays only the odd philosopher believes that the world is made up only of particles. We know that there is no particle without field, that there are fields with no associated particles, and that in addition to physical objects there are chemosystems, biosystems, and sociosystems possessing emergent properties. (See D'Abro (1939) for the decline of mechanism in modern physics, and Vol. 4 for the supraphysical levels.)

Mechanism is dead but *physicalism*, or the program of trying to reduce every high level theory to physics, or perhaps to physics and chemistry, is still very much alive. Physicalism is both an ontology and an epistemology. The ontology of physicalism boils down to the thesis that concrete things differ only in complexity; its epistemology, to the thesis that every property of any whole can be understood in terms of the properties of its components. Such reduction would indicate the direction that the advancement of knowledge must take: "the number of principles needed to explain the totality of phenomena becomes ever smaller (...) Consequently, the number of explanatory principles used may serve as a measure of the level of knowledge attained, the highest being that which gets along with the fewest explanatory principles that are not themselves susceptible of further explanation. Thus the ultimate task of knowing is to make this minimum as small as possible" (Schlick, 1925, p. 13). On this view the progress of knowledge would consist in a progressive reduction of the number of ideas accompanied by a ceaseless multiplication of data.

According to the proponents of the reductionist program, only reduction can attain the unity of science, doubtless a desideratum. Thus the social sciences would be reduced to biology, which would reduce to physics: "each science is a special case of the one that precedes it (...) Biology is not 'just' physics and chemistry, but a very limited, very special and profoundly interesting part of them. So with ecology and sociology" (Medawar, 1974, p. 62). "In its simplest form the *unification program* consists in the unification of all branches of science by means of successive microreductions of the dynamic theories of these branches to one unified theory" (Causey, 1977, pp. 4–5).

How has the reductionist unification program fared? The history of science, technology and the humanities over the past century exhibits not only reductions but also mergers, not only unification but also diversification. The number of theories has become ever greater, and the number of branches of knowledge—even within physics and chemistry—has not ceased to grow. But at the same time the connections among theories and entire fields of research has increased. In short, we are heading towards the unity of knowledge through both reduction and synthesis, and in any case through enrichment not simplification, through interaction, not isolation. Thus we hope soon to be able to understand language not by reducing linguistics to physics or by keeping it detached from all the other branches of learning, but by integrating it with neuroscience, psychology, and social science. (More in Vol. 6.)

Why has the program of unification through reduction met with only limited success, so that we must look for a complementary mechanism—namely integration—to achieve the unity of knowledge? One reason is, no doubt, that the world is far more varied than imagined by physicalism: far from being flat it has a complex level structure: See Figure 13.3. A second reason is that a science is not a theory but a system composed of theories, methods and data, so that, even if the successive microreductions of theories were possible, this would not lead to the reduction of all sciences to one. A third reason is that, in most cases, the reduction of one theory to another is only partial, in that it requires the adjunction of premises not contained in the reducing or basic theory (Ch. 10, Section 3.1). Thus, to account for the very existence of a molecule we need not only quantum mechanics but also a set of subsidiary hypotheses; for example, the theory of the hydrogen molecule is not indebted to that of the hydrogen atom. (More in Vol. 6.)

In addition to the above conceptual obstacles to full reduction there is

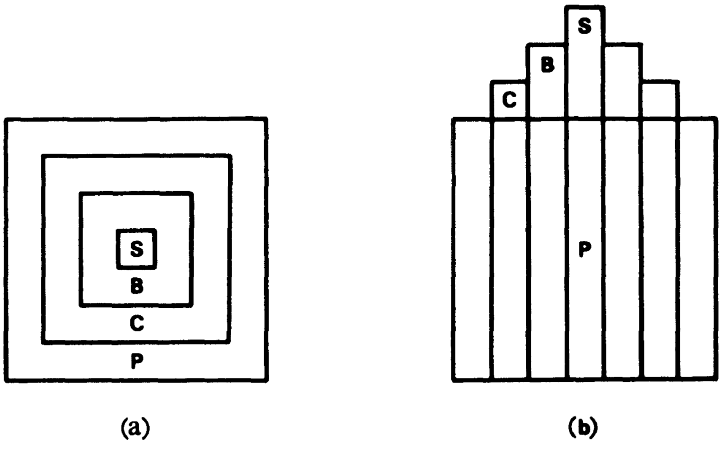


Fig. 13.3. (*a*) Ontological reductionism: reality is a system of Chinese boxes. Higher level things are just special cases of lower levels. (*b*) Ontological emergentism: reality is a telescopic system. Every higher level is composed of things that possess some emergent properties in addition to some of the properties characterizing lower level things. '*P*' stands for the collection of physical things, 'C' for chemosystems, 'B' for biosystems, and 'S' for sociosystems.

one of a practical nature, namely computational difficulty. Take for instance the problem of calculating *ab initio* the possible energy levels of the chromosome of *Escherichia coli*, the best known cell and a comparatively simple one. Since the chromosome consists of a double-stranded DNA molecule with a molecular weight of about 2,800,000,000, in order to account for every single interaction among its subatomic components we would have to set up and solve a wave equation with about 10^8 terms, a task beyond human capacity. It is therefore reasonable to give up such a reductionist attempt and analyze the molecule into manageable modules. In short, reduction cannot always be performed, be it for conceptual or for practical reasons.

Still, the unity of science is a worthy goal, and one that is being attained through both reduction and integration. Why is such unity desirable? First, because the world happens to be one even though immensely rich both quantitatively and qualitatively. This unity suggests that many of the barriers among the various research fields are the product of ignorance or accident, hence unjustified. Such barriers are in our brains not in the external world. Second and consequently, the *rapprochement* and eventual fusion of different fields of inquiry is bound to be fruitful if they study the same objects (the case of ethology and neuroscience) or if one of them studies the components of the systems investigated by the other (the case of psychology and sociology). By getting together with its neighbors, each research field illuminates and checks the other. (Recall the requirement of external consistency or coherence of every body of knowledge with its nearest neighbors: Ch. 12, Section 3.1.)

A fourth reason for wishing to attain or rather strengthen the unity of knowledge is from the limitations of specialization. The division of labor, though necessary to obtain accurate and deep knowledge, may reach a point when it becomes an obstacle to further progress. As Adam Smith noted two centuries ago, narrow specialization blocks technical innovation. The same holds for scientific research: the narrow specialist tends to lose sight of the scientific enterprise as a whole, to the point that he may become an eminent technician ignoring everything that goes on around him. As a result he does not enjoy the benefits of the flow of information and the suggestion of analogies from one field to another, and he is unlikely to be able to cope with new problems requiring an integrated approach. By specializing in excess he has fallen into routine, i.e. the very opposite of innovation. This is one reason that many a scientific, technological and humanistic revolution has been the work of marauders free from the fetters of disciplinary tradition. For example, physiological psychology was founded by two physicists, Helmholtz and Mach; and the Mach bands, which we now attribute to lateral inhibition, were intensively investigated by Békesy, originally a communications engineer. Molecular biology was largely the creation of physicists and chemists: Watson was the only biologist among the fathers of the new discipline. And operations research, or large scale management, resulted from the joint efforts of physicists, mathematicians and engineers—not of administrators or strategists.

We have begun to understand that the current fragmentation of science, technology and the humanities hinders their advancement: that breakthroughs in all fields of inquiry can be obtained only by a judicious combination of specialism with generalism. The model to be imitated is neither the assembly line worker nor the handy man, but instead the specialist with a general background and outlook, who keeps updating and diversifying his knowledge, who keeps abreast of advances in adjoining fields as well as tries his hand at some problems in them, perhaps to the point of changing professions once or twice in life. In short, the best expert is the generalist with specialized knowledge—or, if preferred, the specialist with general interests. It is people of this kind who contribute to creating new research fields or to merging different specialties, and in general help redraw interdisciplinary boundaries. It is also people of this kind whom we seek as our family's doctor: people able to take care of the whole, not by ignoring the findings of the various specialists but by integrating them. If this strategy works to keep us whole, why not use it more often to improve the health of our badly fragmented intellectual culture?

3. Evolution and revolution

3.1. *Paradigm*

Every human being is born into a culture, and every culture includes one or more fields of knowledge, some of which are rather closed belief systems (e.g. religions) whereas others are open research fields (Ch. 2, Section 4.1). Every one of these epistemic fields includes one or more conceptual frameworks. Each such conceptual framework is composed in turn by a general outlook or philosophy, a body of background knowledge, and an accepted thought style which includes certain methods for handling problems of a given type. In mature research fields a few such frameworks dominate at a given moment whereas the competing frameworks are marginal. The dominant conceptual frameworks, or *paragons*, have been variously called *thought styles* (Fleck, 1935) and *paradigms* (Kuhn, 1962). In developing research fields there are no such dominant conceptual frameworks, thought styles, or paradigms. Thus psychology and sociology are still in search of their paradigms, whereas chemistry has its own. (Contrary to widespread opinion, a mature science need not have a single paragon. Thus a chemist may use, in the same research, classical chemical kinetics, a spoke-and-ball model of molecules, and quantum chemistry—three paradigms altogether.)

Those of us who become professional inquirers—scientists, technologists, or humanists—do so by grasping the main features of the paragons in at least one research field, usually in two or more. We learn mostly by studying model cases or *exemplars* (Kuhn's apt term) of problem solving. And we make original contributions when posing or solving new problems within the existing framework, or when introducing some important and viable alterations in the latter. In the first case we conduct, to employ another of Kuhn's favorites, *normal* research. In the second case we engage in *extraordinary* research, which may result in an epistemic breakthrough or even revolution. All this has been known to scientists, technologists, and historians, for quite some time: Kuhn's (1962) modern classic had the merit of bringing it to the fore. What remain problematic are the very notions of a conceptual framework and of a paradigm, and of a revolution in it. Neither of these notions has been elucidated carefully, either by Kuhn or by his followers or critics. (See, e.g. Lakatos and Musgrave, 1970.) Let us apply our tools to clarify those notions.

In Ch. 2, Section 4.1 we characterized a field of inquiry as being formed by a material framework and a conceptual one. The former is constituted by an inquiring community (or rather a community of inquirers), the society supporting (or at least tolerating) it, and the sort of thing that the inquirers study. (In the case of the formal and humanistic disciplines the "things" being studied are conceptual objects, so the expression 'material framework' is somewhat misleading: a better name is wanted.) A *conceptual framework* in any given epistemic field $\mathscr{E}$ was characterized as a septuple $\mathscr{E}_c = \langle G, F, B, P, K, A, M \rangle$, where

G = *general outlook* or philosophical background,
F = *formal background* (logical or mathematical presuppositions),
B = *specific background* (body of borrowed knowledge),
P = *problematics* (collection of problems that may be investigated in $\mathscr{E}$),
K = *fund of knowledge* previously obtained by members of the inquiry community,
A = *objectives* or goals of the inquiry, and
M = *methodics* (collection of methods of $\mathscr{E}$).

Except for the occasional impostor, every member of a research community is engaged in designing or implementing one or more research projects. A *research project* in a research field characterized by a conceptual framework $\mathscr{E}_c = \langle G, F, B, P, K, A, M \rangle$ may be construed as a septuple $\rho = \langle g, f, b, p, k, a, m \rangle$, where every component is a subset of the corresponding component of $\mathscr{E}_c$. Two or more research projects are said to *compete* with one another it they deal with roughly the same problems in different ways, e.g. using different special methods. For example, at one time physicists were divided into corpuscularians and plenists (or continuists); and nowadays sociologists may be classed into holists, individualists, and systemists. There is no competition if the aims are different—e.g. theoretical in one case and practical in the other.

An *exemplar* may be defined as a research project that, (*a*) having proved successful in the past, (*b*) is imitated (taken for a model) for the conduct of further research work. Lakatos (1970) proposed his own notion of a *research programme*, conceived as a sequence of theories in themselves (i.e. dwelling in the platonic realm of ideas), as an "objective reconstruction" of Kuhn's notion of a paradigm. Actually it is an adulteration of it, because for Kuhn (*a*) research is not restricted to theorizing, and (*b*) theories do not hover above social circumstances. The thrust of Kuhn's view is that inquirers do not work in a social vacuum but in a research community. Much the same can be said of the analysis proposed by Stegmüller (1976)—which, moreover, rests on the analysis of theories due to Sneed (1979), criticized in Ch. 9, Section 1.2.

We define *normal research*, be it in science, technology, or the humanities, as the implementation of a research project within an existing conceptual framework and in imitation of some exemplar or paradigm. *Extraordinary research*, on the other hand, is that which may result in a radical innovation in some conceptual framework, such as a substantial change in general outlook, problematics, or methodics. If successful, the new conceptual framework generates new paradigms which inspire a new run of normal research.

Normal research is the bread and butter of investigators, and much of it is exciting. (Even those few revolutionaries who succeed in constructing a new conceptual framework do normal research when they adopt the new paradigm to investigate problems other than those that gave rise to the new framework.) Normal research is often predictable in outline but need not be always so. In fact it sometimes proves our intuitions and expectations wrong, as when Maxwell discovered that the viscosity of a gas is unrelated to its density. (For a good selection of surprises in the course of normal research in contemporary theoretical physics see Peierls (1979).) Moreover most breakthroughs are effected within existing conceptual frameworks. *Examples*: the 18th century work in mathematical analysis (in contrast with the invention of the latter in the previous century); Laplace's work on probability (*vs*. the earlier work); Frege's contribution to logic (*vs*. that of Boole and De Morgan); the axiomatization of set theory (*vs*. its creation by Cantor); modern solid state theory and quantum chemistry (*vs*. quantum mechanics); and even molecular genetics (*vs*. classical genetics) according to Maynard Smith (1972).

Extraordinary research involves a change in thought style and therefore it elicits a reorientation of research. If the outcome is a substantial advance, it constitutes an epistemic revolution—what Bachelard (1938) called a *rupture épistémologique*. Moreover, a successful research project will be said to be an *epistemic revolution*, relative to a given conceptual framework $\mathscr{E}_c$, if and only if (*a*) it involves radical departures in some (not all) of the components of $\mathscr{E}_c$, or (*b*) it opens up a whole new research field (without however cutting ties with all of the existing ones). Clear cases of epistemic revolutions were those effected by Newton, Darwin, Marx, and Cantor. They changed the prevailing modes of thinking in a deep and lasting way.

However, extraordinary research need not be revolutionary: it may ensue in an epistemic counter-revolution, i.e. a partial return to an earlier conceptual framework. (There are never total returns.) More precisely, a research project may be said to be an *epistemic counter-revolution*, relative to a given conceptual framework $\mathscr{E}_c$, if and only if it involves (*a*) giving up, for no good reasons, substantial portions of any of the seven components of $\mathscr{E}_c$, or (*b*) returning to ideas or procedures that had proved to be inadequate in the past and moreover were superseded by $\mathscr{E}_c$. Contemporary cognitivism, with its obsolete mentalism, its disregard for biology, and its lack of concern for experiment, is a rather clear case of counter-revolution; the romantic revolt against positivism (e.g. Feyerabend, 1975) is another. In still other cases extraordinary research results in a mixture of revolution and counter-revolution. An instance of such mixture is behaviorism, with its scrupulous methodics, narrow problematics, and almost total renunciation of theory. Another mixture of revolution and counter-revolution is transformational linguistics, with its formal rigor and exciting new problematics allied to traditional mentalistic psychology and scant concern for empirical test.

Some philosophers, notably Popper (1970), take normal research to be a matter of routine and even dogma, hence full of dangers. This is not so. As we saw a while ago important and surprising breakthroughs can be made within existing conceptual frameworks. Moreover, every epistemic revolution has its roots in some conceptual framework or other. What is true is that normal research is not as glamorous as extraordinary research, and therefore does not often come to the attention of philosophers. And it is also true that clinging to any given conceptual framework, no matter how fruitful it may have been, may end up in dogmatic rigidity: in the refusal to try new theories or proposals, and occasionally also in the refusal to admit defeat in the face of empirical evidence. This point deserves a new paragraph.

Kuhn's most important contributions to methodology are perhaps his ideas that in every science there is a permanent tension between tradition and change, and that negative evidence is treated differently by normal research and by extraordinary research (Kuhn, 1977). The former idea, though obvious, bears hammering, and anyway it is alien to both gradualists (who view the history of knowledge as cumulative) and catastrophists (who focus on revolutions and ignore normal research). The second idea is more original: it is that, whereas normal research attempts to *accommodate* negative evidence to the ruling conceptual framework, extraordinary research uses such anomalies to *undermine* the framework.

The accommodation of negative evidence to the ruling paradigm can be effected, in all honesty, by augmenting the traditional theory with *ad hoc* hypotheses designed to save it, or by proposing new theories conceived in the "spirit" of the prevailing conceptual framework. There is nothing wrong with any of these tactics unless serious anomalies keep cropping up, i.e. unless the prevailing conceptual framework enters in a crisis. In such case it is advisable to try radical alterations. Of course, any such reform projects will be resisted by researchers who have become attached to the old framework. They may become so conservative as to censor the publication of criticism, of new ideas, and even of unfavorable data. But eventually resistance to novelty weakens and the new framework prevails: change is of the essence of science and technology.

3.2. *Revolution*

Human knowledge can advance *gradually*, by *breakthroughs*, or by *revolutions*. Gradual advance consists in accretion or attrition: in gaining some items of information or in discarding others upon being recognized as inadequate. Gradual advance occurs always within some conceptual framework or other. Once in a while a breakthrough occurs within a conceptual framework, namely when an important problem or problem system is solved, so that new problems can be posed within the same framework. And revolutions consist in the emergence of new conceptual frameworks, which replace either old ones or just plain ignorance.

This being so, it is a mistake to opt for either gradualism (favored by empiricism) or catastrophism (favored by both rationalism and irrationalism). The history of knowledge, like that of every other human endeavor and, indeed, any other sector of reality, exhibits not only gradual changes and breakthroughs but also revolutions. The synthesis of gradualism and catastrophism is *evolutionism*. According to this view there is (*a*) *permanence* of some over-arching philosophical principles that propel all objective inquiry (e.g. the theses that reality is lawful and can be known), (*b*) *ceaseless addition and deletion* of data, techniques, hypotheses, and

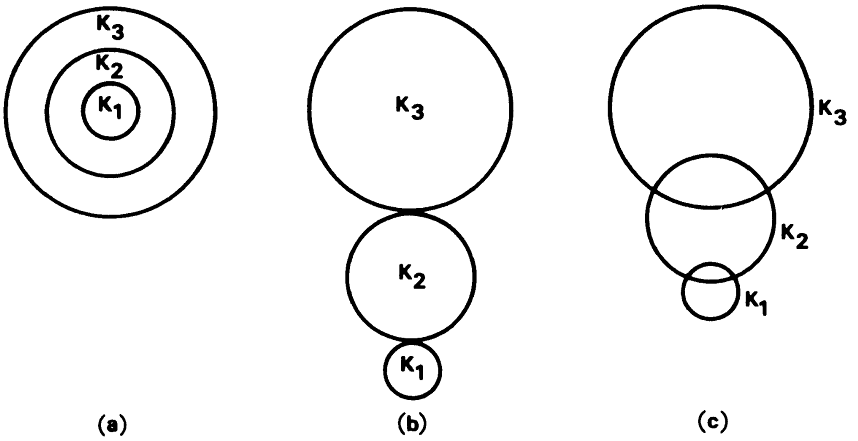


Fig. 13.4. Three models of the evolution of knowledge. (*a*) *Growth*: successive bodies of knowledge include the previous ones. (*b*) *Revolution*: successive bodies of knowledge are mutually incommensurable. (*c*) *Evolution*: successive bodies of knowledge overlap partially.

theories, and (*c*) *occasional revolutions*, using some background knowledge and upsetting other components of it, that result in new conceptual frameworks. See Figure 13.4.

The evolutionary view of the march of knowledge keeps the sound theses of both gradualism and catastrophism while rejecting the false or misleading ones. In particular, evolutionism rejects the fashionable thesis that knowledge advances primarily by replacement, not by addition. This is not always so: mathematical analysis, abstract algebra, genetics, control theory, psychobiology, and economic history—to cite but six scientific revolutions—replaced nothing but ignorance. In these and many other cases there was no rival conceptual framework to be criticized and replaced.

A second, related thesis, is that every epistemic revolution is a response to some crisis. (A research field is said to be in a state of *crisis* if it is stagnant, or is dominated by a single narrow school, or is fragmented into several warrying schools, or is split into many narrow and weakly related specialties, or some of its own accomplishments are threatening its dominant conceptual framework.) True, every epistemic field seems to have passed through a period of crisis, and some fields, such as sociology, seem to be in a permanent state of crisis. However, in some fields breakthroughs and even revolutions may occur without any deep crises. For example the discovery of inconsistencies in mathematical analysis, and later on in set theory, did not throw the mathematical profession into disarray and did not force it to give up any basic principles. The problems were solved with more of the same, i.e. more rigor and further theories.

A third mistaken thesis of catastrophism is that every epistemic revolution razes past accomplishments: that it produces the "collapse" of earlier theories and methods, which are "overthrown" by the victorious rivals. This analogy with politics and warfare is wrong in many cases. Thus Einstein's two relativities, far from demolishing classical physics, constituted its apex: Einstein continued and culminated the work begun by Faraday and Maxwell, as well as by Poisson and Riemann. Furthermore, the news of the demise of classical mechanics and other classical theories is exaggerated, as Mark Twain would say: they are still being worked out and modernized by physicists and mathematicians, and used by engineers. After all, in many cases they help solve problems to an excellent approximation. And in any event even the most drastic revolutions are partial: they alter only some components of the total system of knowledge at any given time.

A fourth mistaken thesis of catastrophism, and a most dangerous one, is that rival frameworks are "incommensurable"—by which "incomparable" seems to be meant, though this is not certain, for romantics are characteristically sloppy. In particular, rival scientific theories would not have a common core of meaning: their senses and reference classes would be entirely disjoint. We have criticized this view in Vol. 1, Ch. 2, Sections 4.2 and 5.1, and Vol. 2, Ch. 7, Section 3.3. Suffice it to note here that, if two theories are deemed to be rival, it is because they share referents and even allow one to pose some common problems. Thus the theory that some mutations are neutral competes with standard genetics, according to which every mutation is either advantageous or, more often, disadvantageous. On the other hand no linguistic theory could possible rival a geological one, for they deal with different domains. In any event it is common practice to compare rival theories and methods, and to choose between them with the help of more or less explicit criteria (Ch. 12). To assert that the choice must be irrational is to distort the picture of actual research and to make an unnecessary concession to irrationalism.

Finally, a fifth and equally dangerous thesis of catastrophism is that each conceptual framework is a sort of mental prison out of which we cannot escape in any rational way: if we do escape it is by an act of faith. Surely this thesis is psychologically and methodologically wrong: scientists and technologists do not operate like mystics or like blind followers of a political ideology, but are often capable of examining their pet theories and methods with a critical eye. They can admit formal or empirical defects and can do something about them, at least tolerate attempts to solve the difficulty. As Popper (1970, p. 56) wrote, "if we try, we can break out of our framework at any time".

In sum, the march of knowledge is smooth in some respects and discontinuous in others. All change, even the most drastic, is partial, not total. (Only crackpots reject the totality of the existing system of knowledge.) And epistemic changes are uneven: at each period some branches of knwledge advance faster than others, thus providing inspiration or even leadership to the less developed ones. Moreover the advancing edge does not fill all the gaps in knowledge: countless pockets of unsolved problems remain behind, some of which are taken up later whereas others are forgotten.

Some of the crucial advances in knowledge have been both philosophical and scientific. In particular, some have consisted in *epistemological revolutions*, or radical changes in the way of gaining knowledge. One of them was introduced by the Ionian philosophers of nature, in particular the atomists: it consisted in replacing myth (in particular religious myth) with naturalistic speculation consistent with (though not suggested by) observation. Another was Archimedes', namely the combination of mathematics with observation and measurement (and possibly also with some experiment) to produce the first mathematical models of things, namely statics and hydrostatics. This process was quickly halted but was resumed sixteen centuries later by Galileo. He replaced the reading of the Scriptures with the "reading the book of Nature" with the help of mathematics, measurement, and experiment. A fourth epistemological revolution was brought about by Newton, who created the first general, exact, and nearly true factual theory in history, and showed how to use mathematics to build theories and formulate problems. A fifth epistemological revolution accompanied the Industrial Revolution and is still going on: it was the birth of science-based technology.

A sixth epistemological revolution started nearly two centuries ago: it consists in studying human nature and human society the same way nature is studied, namely by observation, measurement, experiment, and mathematical modeling. A seventh epistemological revolution was the invention of the theory of evolution by natural selection: from Darwin on not only organisms but also ecosystems, societies, institutions and even molecules are "seen in an evolutionary perspective", i.e. as members of evolutionary trees and as the results of variation and selection. An eighth epistemological revolution is the computer revolution, which makes it possible, among other things, to solve computation problems that are far beyond the reach of the individual brain assisted only by paper and pencil. Finally, an ongoing epistemological revolution is that introduced by physiological psychology, according to which cognition is not an operation of an immaterial mind but a brain process. All these epistemological revolutions shattered long-established traditions, such as those of relying on myth, or on observation unaccompanied by testable theory, or on routine techniques handed down by the craftsman to his apprentice. However, in every case some regulating principles and some bodies of knowledge were preserved. Even the boldest of explorers needs some equipment, although he may have to alter or discard some of it in the course of his travels.

Sometimes three other events are cited as epistemological revolutions: the practice of systematic observation at the beginning of the Modern period, the requirement of operational definitions, and the introduction of considerations on the observer in the quantum theory. Actually neither of these was an epistemological revolution. Man has always observed, and many people have made extensive, meticulous and systematic observations before—though with little result because of lack of good guiding hypotheses. The novelty introduced by the 17th century in this respect was the use of observation (and measurement and experiment) to *check* hypotheses. As for operationism, or the requirement to "define" every concept in terms of instrumental operations, Bridgman (1927, p. 2) regarded it as a guarantee *against* any further revolutions. He thought that we would stop drifting if only we were to anchor every concept to the supposedly solid rock of laboratory operations. Finally, although Bohr, Heisenberg and others did think that they were producing a revolution by invoking the subject every time they interpreted a quantum-theoretical formula in terms of observations, actually they retreated from the ideal of objectivity, so the Copenhagen interpretation they built was *counter*-revolutionary. Fortunately it can be shown (Bunge, 1967c, 1973a) that such interpretation is groundless, for it matches neither the mathematical formalism of the theory nor the actual practice of physicists.

Finally, note that many a revolution in knowledge has been the product neither of new empirical evidence nor of rational criticism. For example, demonology was not discredited by new, adverse empirical evidence, for every bit of experience could be absorbed or "rationalized" by demonology itself. Nor was it discredited by the criticism practised by a few bold and enlightened humanists and theologians, for they had nothing positive to offer instead of a body of belief that could encompass such wide experience.

As Trevor-Roper (1967) has shown, demonology died together with the whole medieval world view, which fell when the new positive philosophy of nature proved more powerful than the old one: it was one of the casualties of the triumph of the new mechanical world view built by Descartes, Galileo, and a few others. Moral: Don't neglect world views, particularly since a theory or practice that is upheld by the prevalent world view is nearly unassailable, whereas one that lacks such support is unlikely to succeed.

4. Limits and prospects

4.1. *Limits to Inquiry*

Are there any limits to what we may know? Barring the skeptical answer—that we can know nothing—three replies to that question have been given. One is optimistic: there are no bounds on inquiry. Another is pessimistic: there are severe limitations on what we can know, and we may already have reached the end of our tether. The third is realistic: there are limitations, yet boundless progress is possible, though not inevitable.

The optimistic thesis is that of the Enlightenment. (See, e.g. the reception speech of the marquis de Condorcet at the French Academy in 1782.) This thesis is little more than an article of faith suggested by the scientific, technological and humanistic progress since about 1600 in Western Europe. The pessimistic thesis is based on certain beliefs on the human mind and society. One of them is that the human mind is a finite capacity information processor composed of a large but finite number of finite elements, so that sooner or later it will exhaust all their possible combinations. Another is nativism, according to which we never learn anything, all our ideas being innate. A third belief is phenomenalism, or the thesis that we can know only appearances and the relations among them, never reality, let alone any essential properties of it. A fourth belief is that every society, and perhaps even the whole of mankind, is bound to decline, for having exhausted its drive, or for being unable to surmount its inner conflicts, or for having degraded its natural resources.

The optimistic thesis ignores the real constraints on inquiry, such as the upper bound on the velocity of propagation of all signals, whereas the pessimistic thesis exaggerates them. The former does not account for any setbacks, such as those of the Counter Reformation, Nazism, Stalinism, and the current restrictions on research budgets in the USA and elsewhere.

As for the pessimistic thesis, recall the fate of all the gloomy prophecies on the future of inquiry. On the eve of the relativistic and atomic revolutions Lord Kelvin prophesied that the next advances in physics would only add a few decimal figures to the value of the physical constants. (For more recent pronouncements on the near completion of knowledge see Schlegel (1967) and Stent (1978).) At the end of the first World War Oswald Spengler wrote the obituary of the so-called Western civilization, and two decades later Cristopher Cauldwell that of "bourgeois" culture. Both prophecies were followed by unprecedented advances in all fields of research. Better refrain from making prophecies and contribute instead to the pursuit of knowledge.

Given the failure of both the optimistic and the pessimistic theses, let us turn to the realistic one. (For details see Bunge (1978a).) There can be no doubt that the pursuit of knowledge is subject to constraints of various kinds: physical, biological, and social. The former consist in the impossibility, be it in principle or in practice, of obtaining certain informations, e.g. about the documents contained in the archives burned during the Thirty Years War. The biological constraints on our cognitive activities boil down to this: we cannot outfox our own brains—and human brains, though marvelously competent, presumably are not the last evolutionary word. And the social limitations on inquiry stem from the economic, political and cultural matrices of every scientific community. The social pressures on a community of inquirers may become so intense that research may stagnate, decay, or even disappear altogether at least for a while. Such social constraints have nothing to do with the physical or biological ones, hence they can be relaxed. Let us take a closer look at the three types of constraint.

The *physical limits* on inquiry consist essentially in that not every desirable item of information is accessible. There are at least two classes of events of this kind: those about which all information has been lost, and events about which any information will be too late. A clear case of lost information is this: if the "big bang" cosmological theory is true, then the history of the universe before the beginning of the expansion has left hardly any traces, so we cannot possibly learn anything about that segment of eternity. A similar case, on a far more modest scale, is the loss of geophysical—in particular geological—information as a result of physical processes such as the melting of rocks. The destruction of prehistorical evidence and historical records is similar. However, we may take heart on recalling that we keep discovering or inventing new ways of finding and interpreting traces of past events. So, the historical record, though hopelessly incomplete, can be reconstructed in part with some imagination.

In other cases there are events but there is no information to be had about them. For example, we cannot know what is happening right now in some remote corner of our galaxy, let alone in other galaxies. If and when the light signals accompanying some such events reach the present abode of mankind, they will give only a partial picture and one that may find no beholders. (In other words, the events about which you can get information are those lying within your own light cone.) However, what a single observer cannot attain, a few generations of observers, bound by a common scientific tradition, may attain: some of the signals that are now leaving more remote places will reach some of our descendants, so they may be able to know what is going on out there.

As for the future, it may be conjectured with the help of laws and a knowledge of present circumstances. Surely some of our predictions are probabilistic, but this is no limitation if the processes themselves are stochastic, as is the case with atomic, genetic, and social processes. For example, although we cannot predict when a particular radioactive atom will decay, we can predict when the radioactivity of a large aggregate of atoms of a certain kind will be halved.

In sum, scientific inquiry is subject to certain physical constraints. These constraints limit the amount of information that can be won, but they need not slow down the pace of scientific progress. We can know more and more about certain things while ignoring everything about others. We need not worry about all we shall never know provided we keep on finding out some of the infinitely many things that can be known.

Let us now turn to the *biological constraints* on inquiry. Ours is the most plastic of all known brains, and so the one capable of learning the most. As society becomes more complex and culture richer, our brains are subjected to an increasing amount and variety of stimuli, which elicit an ever increasing number and variety of synaptic connections. The development of a child's brain is probably far swifter nowadays than it was one century ago. And our readiness to expect novelty and to seek it is certainly greater than that of our ancestors, most of whom resisted novelty. Surely there is more to be learned today, but we do not have to learn everything our ancestors knew. We can learn to pilot airplanes without first having to learn to drive oxcarts. The same holds for science and technology learning: most of the theories and data a student is supposed to learn are comparatively new. Progress makes up for the limited human life span.

Much more serious a limitation on the human cognitive ability would seem to be our limited channel and storage capacity—to say it in computerese. But these are not really disadvantages: if we were capable of admitting all stimuli, we would be unable to concentrate; and if we were capable of recalling every single experience we would hardly be able to conceive of new ideas. Blessed be our sensory inhibitions and memory limitations, for they allow us better to create.

Anyway, the brain-computer analogy is superficial. Human brains are biosystems not artifacts: they develop and, collectively, they evolve. Moreover, unlike other organs, the human associative cortex is highly plastic: it does not deliver the same outputs every time it is presented with the same inputs, and it is in permanent activity. Furthermore it is possible that the various mental functions are performed not by fixed neuronal systems but by itinerant ones formed afresh every time a new task arises (Bindra, 1976). If this is so, even though the number of neurons is finite, that of neuronal systems capable of performing mental functions is practically infinite.

The real limitation of the human brain lies elsewhere and it is not to be found out by comparison with computers. It derives from the inability of our associative cortex to function properly unless it is connected to the brain stem, the hypothalamus, the endocrine system, the peripheral sensors, and, indeed, the whole rest of the body. It is these extracortical systems that keep the associative cortex awake and active—and which at the same time cause the nonrational streaks of human thought. We cannot think unless we are alert and motivated, but if we are alert we cannot avoid some distraction; and if we are motivated we cannot keep totally cool. In sum, pure reason is biologically impossible.

However, we often do manage to think rationally and to make rational decisions, and this because we are not isolated. Sensory deprivation leads to hallucination, and social deprivation to mystical visions. I cannot think rationally all the time, nor can my partner, let alone my rival; but they correct me when I lapse, and all three together manage to set up—unwittingly to be sure—a self-correcting system within which we can be sane and productive, outside of which insane or barren. Human knowledge, particularly of the scientific or technological kind, is produced by individuals in society, and ought therefore to be always public property.

In short, the biological constraints on cognition, though real, are less formidable than they appear at first blush. First, because the human brain is marvelously plastic and capable of spinning out new ideas without end.

Second, because the scientist, technologist or humanist need not, nay must not, rely exclusively on himself. He counts on all his peers, this being why he writes *we* know where the artist writes *I* know. His personal limitations can be overcome by learning from the dead and cooperating with the living. He works, in short, in an inquiring community held together by an information network. (Recall Ch. 3.) Which brings us to the third and last type of constraint on inquiry.

Every inquiring community is a subsystem of some culture or other, which in turn is a subsystem of some society, in turn a component of the international system. The culture of a society is only one of its three artificial subsystems, the other two being its economy and its polity. Each of these subsystems interacts strongly with the other two. (cf. Vol. 4, Ch. 5.) In particular culture—and so science, technology, and the humanities—is subject to intense economic and political pressures. So, far from being autonomous, science, technology and the humanities thrive or wither along with society.

A peasant economy cannot support a space research program; a totalitarian state does not tolerate free inquiry into political problems; and a religion-oriented culture does not encourage studies on the origin of life, biological evolution, the physiology of thinking, or the socioeconomic roots of religion. It does not follow that an industrialized society, with a democratic polity and a lay culture, will necessarily support scientific and humanistic research. It will, provided the prevailing ideology happens to be favorable to inquiry—otherwise not. Nor does it follow that an underdeveloped economy, allied to an authoritarian polity and a backward culture, will necessarily ban all inquiry. It may support some research, sometimes even at great sacrifice, as long as its ideology is friendly towards learning, or for believing mistakenly that basic research delivers technology. In sum, if we wish to understand the social control mechanisms of inquiry we must reckon with ideology, for ideology shapes public opinion, which in turn determines cultural policies, which regulate the only two contributions a society can make to its inquiring communities, namely human and material resources. (Tradition can be borrowed up to a point.)

Science and technology are so much taken for granted nowadays that we tend to forget how new they are and how exceptional were the circumstances of their emergence. Thus it cannot be a coincidence that science and mathematics, as well as philosophy, were invented in Greece during the 5th century B.C. Other civilizations had been more advanced in other respects and they shared a number of basic social and economic traits with the Greek city states, among them slavery. The one trait they did not share was political democracy, which was genuine as far as the Greek citizens were concerned. And political democracy involved civil rights, among them the rights to inquire, question, dissent and argue about everything except religion. Apparently nothing more than this freedom and some leisure was needed for mathematics, science and philosophy to flourish, as inquiry comes spontaneously to man, and social conditions can only stimulate or inhibit it. Technology is another matter: it also requires an industry capable of using it, and this was nonexistent in Antiquity. So, technology proper, i.e. technical knowledge based on science, could not emerge until the mid-seventeenth century. Since then science, technology and industry have been supporting one another, constituting a unique system unknown in previous times. We shall come back to this in Ch. 14, Sections 2.2 and 3.1.

To sum up, although there are physical and biological limits to what men can know, they need not impede scientific, technological or humanistic progress. The collection of knowable facts is a nondenumerable subset of the total set of facts, and there is no limit on the variety of plastic neuronal systems capable of handling those facts. The really important constraints upon the evolution of science, technology, and the humanities, are social—economic, political, and ideological. Hence the importance, for all inquiring communities, of a sociology and politology of knowledge capable of revealing the external stimuli and inhibitions to inquiry. We shall return to this point in Ch. 15.

4.2. *Prospects of Inquiry*

What are the prospects of inquiry in science, technology, and the humanities inspired in science? That is, what is the future of our science-oriented culture? Let us not ask *what* discoveries and inventions will be made, for to do so would be to indulge in fiction—except of course in the relatively uninteresting cases where the outcome of current research is predictable. Let us ask instead the far more radical, important, and topical question *whether* research will be conducted at all in the near future.

Most people seem to be optimistic about the prospects of inquiry, and they offer the following reasons for such attitude. One is that modern civilization needs science, technology, and even the humanities. Indeed, there is no technological advance without research, and much of technological research is indebted to science. It is a question not just of continuing to borrow from past achievements, for many of them become obsolete and also because technology and the economy are constantly running up against new problems calling for new theories, methods, and data. So, if modern civilization—i.e. industrial society in any of its versions—is to go on, research must go ahead. This is the *argument from the practical value* of science. It is persuasive provided one accepts the antecedent, i.e. the thesis that industrial society is worth preserving—not so otherwise.

A second reason given is this. Modern society is becoming increasingly a service society: it needs physicians and dentists, veterinarians and agronomists, electricians and electronics experts, accountants and programmers, and so on and so forth, in increasing numbers relative to workmen and farmers. And those technicians cannot be properly trained, or contribute to the improvement of their techniques, unless they master the fundamentals of science. This is the *argument from services*, which is really complementary to the previous argument. Again, it holds water provided one cares for the service society.

A third argument is that science and technology have become no less than the left and right cerebral hemispheres of modern culture. They are intrinsically valuable in addition to having a practical value. Our schools teach them, and the public demands scientific and technological literature as well as gadgets. This is the *argument from the cultural value* of technology. It is correct but it has little practical value, for it will not persuade anyone who is mainly interested in fast profit or power. Besides, science and technology, unlike their industrial outputs, are still elitist. Only a very small fraction of the total population is ever given the opportunity of experiencing the thrill of discovery or invention. Most of us, even most philosophers of science and technology, have never done any research in either science or technology, and so have as much experience of them as the deaf have of music or the blind of painting. Worse, the antiscientific and pseudoscientific publications have a far wider readership than the scientific and parascientific ones.

Finally there is the *argument from politics*, according to which scientists, technologists, and their numerous auxiliaries—technicians, librarians, etc.—have become a substantial sector of modern society, so their needs and aspirations can no longer be safely ignored by the politicians. This argument carries little weight because most scientists and technologists are politically naive and passive: they have no powerful lobby in any congress.

All four arguments come down to this: Modern civilization, West and East, involves science and technology; hence, if it is to continue, science and technology have got to have a future. Unassailable. But who assures us that modern civilization will continue? There is no guarantee: no law of continuity, let alone progress. On the contrary, there are many disturbing signs of decline. Let us review them quickly.

A handful of recent events has badly shaken our faith in progress, particularly in the continued march forward of knowledge. One is the accelerated pace of the arms race, which employs roughly half of the best engineers of the world, wastes nearly two thousand million dollars per day, and is likely to end up in the ultimate democracy of ashes. The second is the rapid depletion of non-renewable natural resources amidst the nearly total indifference of most governments. The third is the near-sighted science and humanities policy adopted by a number of governments, that have shifted the bulk of public support from basic to applied science and technology. The fourth is the spread of religious fundamentalism and authoritarian ideologies, their attack on basic scientific theories such as the theory of evolution, and their attempt to restrict the right to know. The fifth is the revolt against reason, and in particular against science and technology, that is spreading among well-intentioned but ignorant young people who, mistakenly, see in inquiry an enemy of the people.

For all of these reasons we should not take science and technology for granted. Instead, we should ponder seriously the initial question, which might not have been asked in 1780, 1880, or 1980: Do science and technology, and indeed modern civilization, have a future at all? It is not that science and technology are in the grips of a conceptual or methodological crisis. On the contrary, the prospects for further sensational (big and unexpected) breakthroughs in every field of research are brighter than ever. But science and technology are the work of human beings who need a favorable environment. If such an environment is not provided, the coming generations, if any, may lose interest in science and technology. If this happens our offspring are in for a new Dark Age. We may already have entered a period of scientific and technological decline. The decline, if there is one, may be temporary or terminal. It is up to us which it is to be. There are no iron or even tin laws of scientific and technological development. Science and technology will become what scientists, technologists and their fellow human beings choose.

5. CONCLUDING REMARKS

Knowledge cannot be stored permanently, either in books or in a spiritual world, because it ages. Sooner or later every bit of knowledge becomes stale or, if still valuable, it is bound to be incorporated into new bodies of knowledge, where it acquires a new value. Research takes care of such epistemic metabolism.

Knowledge may pass away, inquiry need not—that is, provided there is the will to find out. Science, technology and the modern humanities are characterized not so much by the amount of knowledge they have amassed, as by their research pace. A swift pace, more than any other characteristic, is what separates them from other forms of knowledge.

Knowledge evolves both gradually and by fits. It evolves as research of all kinds—mathematical, experimental, historical, etc.—yields new results. Some of the results of research fill a predesigned mold—conceptual framework—whereas others break it. In either case good reasons are given for any departure from tradition. And no such departure is a total break with the past: there is discontinuity in some respects but continuity in others.

In principle the progress of knowledge is boundless, but in practice there are restrictions. For example, in principle every single fossil in North America could be dug out; but such digging on a continental scale is practically impossible. In general, there is a gap between what can be known in principle and what can be known in practice.

It is likely that what remains to be known is infinitely greater than what we already know. Yet we do possess already infinite knowledge of sorts. Thus we know how to construct infinite sets, such as continuous functions, and we know law statements that cover infinitely many states, such as the formulas for the propagation of a wave and for the energy levels of an atom.

The enthusiasts as well as the enemies of research and development tend to believe that science and technology cannot be stopped: that, once launched, they are bound to move forward. Taken literally this belief is absurd, for it is inquirers who discover and invent, compute and design, criticize and evaluate—and researchers may cease to exist either for lack of motivation or as a result of a nuclear holocaust. After all, inquiry has declined dramatically several times before and seems to be declining right now: there is nothing necessary about the progress of knowledge. What is true is that creative scientists and technologists, as well as humanists, are mainly inner-directed: they cannot help posing problems, making observations, framing hypotheses, or drawing blueprints—that is, provided they are tolerated by society. And this is precisely the point: society, or rather its rulers, may not cherish research, at least of certain kinds, for regarding it as dispensable luxury or as subversive.

In sum, mutability is one of the trademarks of scientific and technological knowledge. But it is not the only one, and in any case there must be differences between science and technology. Let us investigate them.

CHAPTER 14

KINDS OF KNOWLEDGE

What kinds of knowledge are there, what characterizes every one of them, and how are they related? These are some of the problems to be investigated in the present chapter. They are not very old: they were born with secular thinking, which in the West usually dates back to the Ionian philosophers of nature, and they were not investigated vigorously until the Renaissance. Thus the distinction between mathematics and the other sciences is attributed to Leibniz, and that between philosophy and the sciences to Kant. And the distinction among the various sciences was not drawn until the beginning of the 19th century, when the first modern universities were organized.

It is common to distinguish the following ten knowledge genera: ordinary knowledge, technical prescientific knowledge (technics), basic (formal and factual) science, applied science, technology, the humanities, the arts, ideologies, religions, and pseudoscience. However, the criteria for such distinctions are seldom explicit and clear. Nor is there consensus about such criteria. Thus according to some, the fields of knowledge differ by their subject matter; to others, by their method; to still others, by the kind of service they render to society; and to some, by the way their workers are trained and communicate among themselves. Finally, some scholars hold that such distinctions are conventional—human knowledge being in one piece—or pointless, human knowledge being non-existing.

We shall propose explicit definitions of the genera of specialized knowledge. However, we shall also note their interrelations for, in one sense, human knowledge is in one piece. The interrelations are of two kinds: systematic and heuristic. Two fields of knowledge can 'be said to be *systematically* related to one another if at least one of them borrows some knowledge from the other. Thus biology and chemistry are systematically related to each other, and so are economics and mathematics. And two fields of knowledge will be said to be *heuristically* related to one another if work in one of them is helped or hindered by the other. Thus religion discourages scientific inquiry, particularly into its own dogmas, but it may occasionally help some investigators conceive fruitful speculations concerning unobservable entities or properties. (However, religion is primarily a matter of faith not knowledge, so it is not really a field of knowledge.)

Since every field of knowledge is related both systematically and heuristically to some other field, one can rightfully speak of the unity of human knowledge amidst its amazing diversity. In fact we shall speak of the *system* of human knowledge—both as a system of knowing subjects and as a conceptual system. Moreover, since the production of knowledge is intimately related to the production, use and exchange of other artifacts, we may also speak of the *total system* of production and exchange of goods and services of all kinds, from theories, blueprints and works of art to cultivars, machines and educational services.

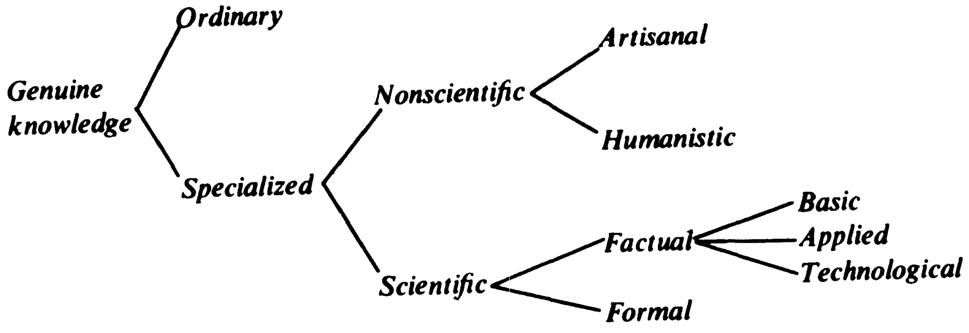
1. FIELDS OF KNOWLEDGE

1.1. *Knowledge Genera*

There are many ways of classing items, in particular bits of knowledge: as many as criteria (or equivalence relations) are adopted. (See Ch. 9, Section 1.1.) Indeed one may divide knowledge into genuine and bogus, theoretical and practical, scientific and nonscientific, factual and nonfactual, and so on. Which classing is the most convenient depends upon our goal.

From an epistemological viewpoint the first division, between genuine and bogus knowledge, is the most important of all. We define *genuine knowledge* as knowledge that is at least partially true, and *bogus knowledge* as thoroughly false knowledge—false either because it refers to nonexistents or because it represents existents in an utterly false manner. However, the frontiers between genuine and bogus knowledge are not fixed. In particular, some genuine knowledge may come to be recognized as bogus. But the converse process is exceptional: we have here the analog of Gresham's law for currency. More on bogus or illusory knowledge in Section 4.

From an epistemological point of view the following classification of kinds of genuine knowledge is of interest:



This schema must be accompanied by a number of qualifications. First, the bits of knowledge that compose every genus or sector, such as ordinary knowledge, vary with place and time. Thus, the basics of arithmetic, physics and technology have become common knowledge throughout the industrialized world, but are still specialized in the underdeveloped nations. Second, not all ordinary knowledge is widespread although it is in principle accessible to everyone. In fact some of it is expressed by works of art such as novels, dramas, movies, paintings, and sculptures. (Shakespeare and Molière, Balzac and Tolstoy, Chaplin and Costa Gavras have taught us more about ourselves than the whole of behaviorist psychology put together. On the other hand music, ballet and non-representational painting are hardly types of knowledge.) Third, specialized knowledge is not all special or narrow: some of it, such as mathematics, epistemology and ontology, is extremely general. Fourth, what is scientific today may have started by being nonscientific; psychology and social history are cases in point. Fifth, what starts as basic may end up by becoming applied or vice versa. Thus solid state physics, once a branch of basic physics, is now part of applied physics. Sixth, the usual classification of disciplines need not fit the above schema: indeed some of them—e.g. forestry, medicine and psychology—have basic, applied, and technological aspects. We shall deal with such interesting mongrels in Section 3.2.

What about value theory, aesthetics, ethics, and ideology: do they qualify as types of knowledge, and if so which place do they occupy in our schema? *Prima facie* neither does, because they are all concerned with valuations, and these are not cognitive acts although they may be performed in the light of knowledge. However, not every statement in value theory, aesthetics, ethics or ideology is a value judgment: there are also strictly descriptive or explanatory statements in those fields. In particular, a value theory is supposed to encapsulate our knowledge about values, and an ethical theory is supposed to account for the nature and role of behavior codes. So, those fields do have a cognitive content.

Moreover, not every value judgment is unjustifiable or irrational. Thus we may have objective grounds for preferring one kind of food to another (for a given end and in given circumstances), or one method of calculation or measurement to another, or one social program to another (or to none). In particular ideology, though committed to values, may be firmly based on knowledge. Thus suppose that most people in a given community live below the poverty line (first factual premise). Suppose also that we discover that this situation is potentially explosive (second factual premise). It

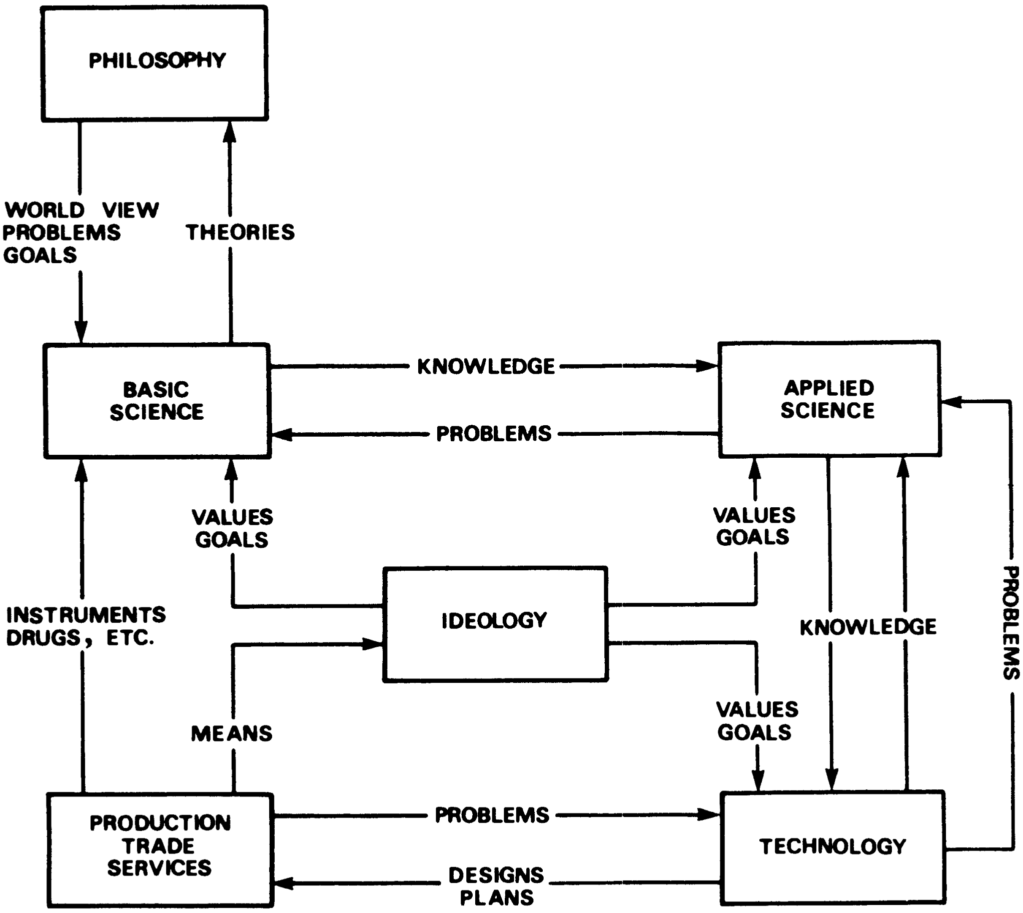

PHILOSOPHY
WORLD VIEW
PROBLEMS
GOALS
THEORIES
BASIC SCIENCE
KNOWLEDGE
PROBLEMS
APPLIED SCIENCE
VALUES
GOALS
VALUES
GOALS
INSTRUMENTS
DRUGS, ETC.
IDEOLOGY
KNOWLEDGE
PROBLEMS
MEANS
VALUES
GOALS
PRODUCTION
TRADE
SERVICES
PROBLEMS
DESIGNS
PLANS
TECHNOLOGY

Fig. 14.1. The system of production and exchange of scientific and technological knowledge, artifacts and services in an industrialized society.

follows that it is desirable both practically and morally to set up a social program to remedy the situation—e.g. to change the distribution of wealth, or introduce new jobs, or offer technical training, or what have you. Such mixture of knowledge and valuation does not differ in principle from the decision to replace one method of measurement with a second one given that the latter is more precise, and that greater accuracy is desirable. In sum, although the value disciplines and the ideologies are not properly included in knowledge they overlap with it. Therefore they must be shown in any diagram representing the system of knowledge: Figure 14.1.

1.2. *Research Field*

In Ch. 2, Section 4.1 we elucidated the notion of an epistemic field, and distinguished two kinds of epistemic fields: the belief system and the field of inquiry or research field. We shall now take a closer look at the latter. We start by defining a *family of research fields* as a set every member $\mathcal{R}$ of which is representable by a 10-tuple

$$\mathcal{R} = \langle C, S, D, G, F, B, P, K, A, M \rangle,$$

where, at any given moment,

(i) C, the *research community* of $\mathcal{R}$, is a system composed of persons who have received a specialized training, hold strong information links amongst them, and initiate or continue a tradition of inquiry;

(ii) S is the *society* (complete with its culture, economy, and polity) that hosts C and encourages or at least tolerates the activities of the components of C;

(iii) D, the *domain* or *universe of discourse* of $\mathcal{R}$, is the collection of objects of study of $\mathcal{R}$;

(iv) G, the *general outlook* or *philosophical background* of $\mathcal{R}$, is composed of ontological theses (concerning the nature of the D's), epistemological principles (about the nature of inquiry into the D's), and ethical rules (about the proper behavior of the inquirers in C);

(v) F, the *formal background* of $\mathcal{R}$, is the collection of logical and mathematical theories that are or can be used by members of C, in studying the D's;

(vi) B, the *specific background* of $\mathcal{R}$, is the collection of items of knowledge obtained in other fields of inquiry and utilizable by the C's in studying D's;

(vii) P, the *problematics* of $\mathcal{R}$, is the collection of problems (actual or potential) that can be investigated by the C's;

(viii) K, the *fund of knowledge* of $\mathcal{R}$, is the collection of items of knowledge utilized by C and obtained by it at previous times;

(ix) A is the set of *aims* or *goals* of the members of C with regard to their study of D's;

(x) M, the *methodics* (usually misnamed 'methodology') of $\mathcal{R}$, is the collection of methods utilizable by members of C in their study of D's.

(xi) There is at least one other (*contiguous*) research field $\mathcal{R}'$ in the same family of fields of inquiry, such that (*a*) $\mathcal{R}$ *and* $\mathcal{R}'$ share some items in their general outlooks, formal backgrounds, specific backgrounds, funds of knowledge, aims, and methodics; (*b*) either the domain of one of the two fields, $\mathcal{R}$ and $\mathcal{R}'$, is included in that of the other, or each member of the domain of one of the fields is a component of a system in the domain of the other.

(xii) The membership of every one of the last eight components of $\mathscr{R}$ *changes*, however slowly, *as a result of inquiry* in the same field or in related fields.

Note the following points. First, only the first two components of the ten-tuple, i.e. C and S, are individuals and, more precisely, concrete systems; the remaining components are collections (variable in time). The point of listing C and S explicitly is to remind ourselves that knowledge is not self-existing but an activity performed by real people in a concrete social environment. So much so that a drastic change in either C (such as the appearance of a couple of geniuses or of a new mode of communication) or S (such as a sudden liberalization or a dictatorship) may alter radically the entire field of inquiry. Of course C is a subsystem of S and therefore a description of C should be included in any thorough description of S. However, C should be listed separately because it may emerge or become extinct even if the society hosting it goes on roughly as before at least for a while.

Second, listing the *domain* D or collection of objects of study of a research field is necessary to characterize it—*pace* Kuhn (1962)—because its problematics, hence its methodics, depends largely upon the nature of its D's. Thus $\mathscr{R}$ is a *factual* field only if some of its D's are concrete things. However, it is not possible to identify the fields of inquiry exclusively by their subject matter, because a number of disciplines may share the same referents: think of human genetics and physiology, anthropology and history. So, mention of D, though necessary, is insufficient for an unambiguous characterization of $\mathscr{R}$: it must be supplemented with a characterization of the problematics, the methodics, and the aims.

Third, the *general outlook* G is essential for, as Myrdal (1969, p. 51) said, before there can be a view there must be a viewpoint. Thus the religious viewpoint is composed of an ontology, an epistemology and an ethics radically different from the scientific viewpoint. (More in Section 4.2.)

Fourth, the *formal background* F is the collection of logical and mathematical theories that members of C may use, whether explicitly or tacitly. True, most scientists and technologists care little for formal logic and for the theorems proving the rules of calculation they employ. Still, once in a while they are forced to resort to them, and they are always supposed to admit corrections to their invalid reasonings and calculations.

Fifth, explicit mention of the *specific background* B of a research field is methodologically and practically important, for it places the burden of some proofs on a different research community and it indicates how well integrated the given research field is with the other members of the same family. Thus a chemist normally relies on results obtained by physicists. On the other hand pseudoscientists make no use of knowledge obtained in other epistemic fields.

Sixth, the *problematics P* of a research field is one of its characteristics, for two different research fields may share everything but their problematics. For example, whereas chemical statics studies the properties of molecules, chemical kinetics and dynamics study the formation and breakdown of molecules (i.e. chemical reactions).

Seventh, the *fund of knowledge K* of a research field is not so much the sum total of results obtained in it to date, as the current stock, i.e. the collection of data, hypotheses, theories, and methods that are still available for further research. K is thus the state of $\mathscr{R}$ at the given time.

Eighth, the *aims or goals A* are equally characteristic of a research field, for one and the same thing may be studied with different aims and hence methods—e.g. to know it better, or else for gaining control over it.

Ninth, the *methodics M* must be listed because a problem may be investigated with radically different methods—e.g. theoretically or empirically. A research field is *theoretical* if its methodics includes only conceptual procedures, and *factual* if it also includes empirical procedures.

Tenth, the existence of *contiguous* research fields (condition xi above) is necessary for a research field, whereas belief systems and pseudosciences are characteristically isolated from other epistemic fields. For example, neither theology nor psychoanalysis have intimate relations with any other epistemic fields.

Eleventh, note that condition (xii) stipulates not only that a research field should change, but also that it should do so *as a result of inquiry* not of mere controversy or in compliance with governmental or ecclesiastical demands.

Let us examine three characteristically modern fields of inquiry: basic science, applied science, and technology.

2. SCIENCE AND TECHNOLOGY

2.1. *Basic Science*

Most philosophers of science believe that science is characterized by a single peculiar trait. In fact the following views on science can be found in the literature:

(*a*) The *consensus* view holds that, whereas theology, pseudoscience and the humanities are rife with controversy, science is uncontroversial or at least aims at "a consensus of rational opinion over the widest possible field" (Ziman, 1979, p. 3). This characterization is inadequate, for (*a*) every field of active basic research teems with controversy, and (*b*) consensus is at most an uncertain indicator of truth, not a goal of research. What is true is that science has means for settling controversies in the long run. It is also true that applied science is far less controversial than basic science, if only because it consists in an application of the firmest results of basic science. Still, the aim of applied research is possibly useful knowledge rather than consensus, which is at best a byproduct.

(*b*) The *empirical content* doctrine maintains that, unlike other types of inquiry, science accepts only empirical data and inductive syntheses thereof, never speculation. Though still popular in the science textbooks, this view was refuted long ago by the very emergence of theoretical physics, theoretical chemistry, theoretical biology, and other fields teeming with concepts representing non-observable entities or properties. Several philosophers, notably Whewell (1847) and Popper (1959), have discredited that doctrine.

(*c*) The *success* view claims that for science only success counts—or, as James put it, that truth is "what it is better for us to believe". This pragmatist view may fit technology, not basic science, which is after truth, depth, and systemicity.

(*d*) The *formalist* doctrine holds that a body of knowledge is scientific only when it has been thoroughly mathematized. This characterization is too wide on the one hand, for it accepts as scientific much stuff that is not—e.g. beautiful mathematical theories of the nonexisting free market in equilibrium. And on the other hand it is too narrow, for it disqualifies experimental science as well as young science, which is often premathematical. The truth is that science cannot advance beyond a certain point without using mathematics.

(*e*) *Refutationism* maintains that the mark of science is refutability, i.e. its dealing exclusively with hypotheses that are conceivably refutable (Popper, 1959). This characterization ignores fact finding research, be it in the laboratory or in the field. And it admits as scientific all refuted beliefs, such as astrology and creationism, whereas it rejects the most general scientific theories, for not being refutable without further ado (i.e. without the addition of subsidiary assumptions).

(*f*) *Methodism* holds that the sole requisite for science is adopting the scientific method. But if every application of the scientific method were indeed a piece of scientific research, then testing the mental ability of atoms, trying to catch ghosts with special nets, or investigating people who claim to be able to read printed matter stuffed in their ears, would pass for scientific research provided certain precautions were observed.

(*g*) *Sociologism* claims that science is what scientists do. This stipulation is inadequate for occasionally scientists indulge in nonscientific activities, such as mindless data collection, untestable speculation, or even straight pseudoscience.

Since none of the above popular characterizations of science work, we must formulate our own. We shall do it by specifying the general definition of a family of research fields proposed in Section 1.2. A *family of scientific research fields* is a set every member $\mathcal{R}$ of which is representable by a 10-tuple

$$\mathcal{R} = \langle C, S, D, G, F, B, P, K, A, M \rangle,$$

where, at any given moment,

(i) the *research community* C has the same general characteristics as those of any other research field;

(ii) the *host society* S of C has the same general characteristics as those of any other research field;

(iii) the *domain* D of $\mathcal{R}$ is composed exclusively of (certified or putatively) *real entities* (rather than, say, freely floating ideas) past, present or future;

(iv) the *general outlook* or *philosophical background* G of $\mathcal{R}$ consists of: (*a*) an *ontology of changing things* (rather than, say, one of ghostly or unchanging entities); (*b*) a *realistic epistemology* (instead of, say, an idealistic or a conventionalist one), and (*c*) the ethos of the *free search for truth, depth, and system* (rather than, say, the ethos of faith or that of the bound quest for utility, profit, power or consensus);

(v) the *formal background* F of $\mathcal{R}$ is a collection of *up to date* logical and mathematical theories (rather than being empty or formed by obsolete formal theories);

(vi) the *specific background* B of $\mathcal{R}$ is a collection of *up to date and reasonably well confirmed* (yet corrigible) data, hypotheses and theories, and of reasonably effective research methods, obtained in other research fields relevant to $\mathcal{R}$;

(vii) the *problematics* P of $\mathcal{R}$ consists exclusively of *cognitive problems* concerning the nature (in particular the laws) of the members of D, as well as problems concerning other components of $\mathcal{R}$;

(viii) the *fund of knowledge* K of $\mathcal{R}$ is a collection *of up to date and testable* (though not final) theories, hypotheses, and data compatible with those in B, and obtained by members of C at previous times;

(ix) the *aims* A of the members of C include *discovering or using the laws* of the D's, systematizing (into theories) hypotheses about D's, and refining methods in M;

(x) the *methodics* M of $\mathscr{R}$ consist exclusively of *scrutable* (checkable, analyzable, criticizable) and *justifiable* (explainable) procedures, in the first place the scientific method;

(xi) there is at least one other *contiguous* scientific research field with the general characteristics noted with reference to research fields in general;

(xii) the membership of every one of the last eight components of $\mathscr{R}$ *changes*, however slowly at times, *as a result of scientific research* in the same field as well as in related fields of scientific inquiry.

Any research field that fails to satisfy even approximately all of the above twelve conditions will be said to be *nonscientific*. A research field that satisfies them approximately may be called a *semiscience* or *protoscience*. And if, in addition, it is evolving towards the full compliance of them all, it may be called an *emerging* or *developing* science. On the other hand any field of knowledge that is nonscientific but is advertised and sold as scientific will be said to be *pseudoscientific* (or a *fake* or *bogus* science). The difference between science and protoscience is a matter of degree, that between science and pseudoscience is one of kind. The difference between protoscience and pseudoscience parallels that between error and deception. Physics has been the paragon of science since Galileo, psychology and sociology are developing sciences, literary criticism is a nonscientific research field, and parapsychology is a bogus science. (See Section 4.1 for a substantiation of this charge.)

Note the following points. First, the research community must be a *system* proper, not an isolated individual or an aggregate of isolated researchers. However, in the absence of a local research community the scientific investigator will belong to a national, regional, or international research community by keeping abreast of the literature and contributing to it. Only crackpots are marginal.

Second, not every society is capable or willing to support a community devoted to scientific research proper. Thus theocratic societies discourage scientific research, and underdeveloped ones lack the material resources to encourage it (although they often find resources to wage war). So, what these societies support or tolerate are small communities of protoscientists, such as naturalists rather than full-fledged biologists. (More on the social conditions of science in Barber (1962) and Bunge (1980b).)

Third, the requirement that D be formed by *real entities* does not exclude speculation on still undiscovered entities. The point is that such speculation should be testable and should eventually yield results one way or another.

Fourth, the importance of the *philosophical background G* of science is often underrated, particularly by those who define science in terms of its method. They are apt to cite the cases of Newton, who was a religious believer, Mach, who held a sensualist ontology and epistemology, and Poincaré, who defended a conventionalist epistemology. One may rejoin that, while doing science, not philosophy, Newton, Mach and Poincaré were being unfaithful to their own philosophies. For example, there is no mention of God in Newton's equations of motion; Mach presumably took it for granted that his instruments did not vanish when he left his laboratory; and Poincaré admitted that Maxwell's equations are law statements not conventions. There is no science without some ontology and some epistemology. To begin with, all the fundamental concepts of science, such as those of thing and property, state and change of state, possibility and actuality, space and time, life and mind, artifact and society, are ontological. Secondly, when exploring some uncharted territory the scientist is tacitly guided by a number of ontological and epistemological principles. For one thing he presupposes that the most general known laws hold in the new territory, and that the most general methodological principles will help him to explore it. (But of course he is ready to correct such assumptions if proved wrong.) If he believes in objective possibility he will investigate things-in-their-environment instead of trying to account for their behavior exclusively in terms of environmental agencies. If he believes in randomness he will try stochastic models, otherwise he will limit himself to "deterministic" ones. If he is an inductivist he will collect as many facts as possible before hazarding any hypotheses. If he is a deductivist he will prefer exploring the logical consequences of hypotheses proposed by others—and so on and so forth. (For more on the metaphysics inherent in science see Whewell (1847), Agassi (1964), Holton (1973), Bunge (1977a).)

The *ethical* component of the general outlook is equally important to tell science from other epistemic fields. The morality of science is largely built into its methodics, this being why we often fail to detect it. According to Merton (1973) this code of behavior includes intellectual honesty, integrity, humility, disinterestedness, organized skepticism, universalism, impersonality, and communism of intellectual property. (Actually scientists are ambitious and often arrogant rather than humble. But they are supposed to be modest in an epistemological sense, i.e. ready to admit ignorance and error.) This code may not be detected as long as scientists go about their business as usual, but it comes to the fore as soon as one of them violates it—e.g. if he doctors data, plagiarizes, distorts results in the service of some ideology, or utilizes them for mercenary or destructive purposes. So much for the philosophy inherent in scientific research—a philosophy without which science cannot be told apart from other fields of knowledge. Let us go on with our remarks on the list of peculiarities of science.

Fifth, the *formal background F* of a science is an (uncertain) indicator of its degree of development. (No wonder that formalists tend to equate scientificity with formalization.) Thus physicists find occasion to use practically all of mathematics. Hence there is no such thing as a selected and fixed collection of theories making up "the mathematics for physicists". The same holds for all the other sciences. If a given branch of mathematics has not yet been used in a certain science, this may indicate only that the workers in it do not know about the existence of that branch or of its possible uses in their science. Hence it is mistaken to teach science students only those mathematical theories that happened to be of use in the past.

Sixth, the *specific background B* of a science is another indicator of its degree of development, as well as an indicator of the ontic level it deals with. Every science presupposes some other sciences. Thus biology presupposes chemistry, which in turn borrows some physics, which uses mathematics, which uses logic. No science borrows all of the knowledge in the sciences it presupposes, but a science with few debts is either very fundamental or very backward. This notion of intellectual debt (or presupposition) suggests the following stratification of the basic or pure sciences:

Formal sciences: mainly logic and mathematics
Physiosciences: physics, astronomy, earth sciences
Chemoscience: chemistry
Biosciences: biophysics, biochemistry (including molecular biology), genetics, physiology, ecology, biogeography, palaeobiology, neuroscience, etc.
Sociosciences: anthropology, sociology, economics, politology, history, etc.

In addition to these pure basic sciences we have a number of mixed or mongrel sciences, such as psychology and linguistics, which are both biological and sociological. (More on mixtures in Section 3.2.) On the other hand physical chemistry, which studies the physical aspects of chemical processes, is a branch of chemistry. Likewise human geography and economic sociology should be counted as parts of sociology.

Seventh, the *problematics P* of a science can be as vast as its workers wish it to be—whence the importance of a liberal education for scientists. In fact we may distinguish eight kinds of problems in each science:

Domain (referent) questions, such as "What is special relativity about: physical entities or measurement operations?", and "What are the referents of social science: individuals or social systems?"

Philosophical problems, such as "Do forms, in particular shapes, have a driving force, and so can they explain morphogenesis (as René Thom believes), or are they rather an outcome of process?", "Even though theory choice is actually done in an intuitive way, would it be possible to set up definite decision rules?", and "Is vivisection morally justified?".

Formal background problems, such as "What do we lose by giving up the excluded middle principle?", and "Does the given equation have any continuous solutions in such and such interval?".

Specific background problems, such as "Is the thermodynamics used by chemists on a sound basis?", and "Does kin selection (assumed by sociobiology) really operate?".

Problematics problems, such as "Is this problem well-conceived?", and "Is that problem worth being investigated?"

Fund of knowledge problems, such as "Are all redshifts of extragalactic objects to be interpreted as indicators of the expansion of the universe?", and "Are all mutations consequential, or are some of them neutral?".

Aim problems, such as "Should psychology explain the mental in neurophysiological terms?", and "Should social science seek societal laws?"

Methodological problems, of the forms "Is this method reliable?", and "How could we go about measuring that property?".

Eighth, the *fund of knowledge K* of a science may be lean or voluminous, depending on the state of the art, but it cannot be empty, for one cannot start from scratch, as the very formulation of a research problem requires some knowledge. Even a new science, when born, possesses some fund of knowledge, either borrowed from ordinary or artisanal knowledge, or inherited from a parent science.

Ninth, the *aims* of a science include the search for or use of laws, since the supreme goal of science is to understand reality, and such understanding can only be obtained with the use of laws. *No laws, no science proper*—at most some protoscience. For example, a scholarly study on the concept of cow in the Rigveda, or on the nature of angels according to the authors of the Old Testament, does not qualify as scientific because it does not seek to find any laws and it does not even make use of any known laws. In general, textual investigation does not qualify as scientific even if it uses scientific means such as chemical analysis or soft X-rays, for its aim is not that of science. The same holds for art history, intellectual history, and other research fields.

Tenth, the *methodics* of any science includes not only its peculiar techniques but also the scientific method (Ch. 7, Section 2.2). A collection of techniques, e.g. for producing high pressures or high vacua, or for measuring the effects of reinforcement on the learning of philosophy, does not constitute a science: methods are means not ends, and they cannot be applied or evaluated apart from a problematics and an aim. Merely exploiting a given technique for obtaining or processing data without any ulterior purposes is not doing science but just keeping busy and possibly salaried.

So much for our characterization of basic science. It may be objected that any set of criteria of scientificity is bound to change in time, so that what is now regarded as basic science may tomorrow be regarded as something else, e.g. applied science or even pseudoscience. (See, e.g. Hyman, 1982.) Granted. But the mutability of criteria does not prove that they do not exist or that they are futile. Likewise the standards of mathematical rigor are changeable in time, sometimes quickly, but this proves only that rigor can be improved, not that it does not exist. We need a definite set of criteria of scientificity, however time-bound, if we wish to produce, teach, and promote genuine basic science rather than something else, be it technology or ideology, and if we do not wish to be taken in by charlatans posing as scientists.

2.2. *Applied Science and Technology*

The expression "applied science" is ambiguous. It may signify the application of one science to another (e.g. of physics to biology) or the use of scientific knowledge to investigate problems the solutions to which may acquire practical importance—e.g. in industry, education, or politics. The application of the probability calculus to statistics, of physics and chemistry to the earth sciences and meteorology, and of biology to medical problems, exemplify the first concept of applied science. Needless to say, this kind of applied science does not follow automatically from the corresponding basic sciences. For example, the earth sciences are not deducible from the sole principles of physics and chemistry. Far from it, they call for specific theoretical models (of, e.g. the earth's interior), specific methods (e.g. for seismological measurements), and specific data (e.g. about the configuration of the terrestrial magnetic field). Moreover the earth sciences involve specific concepts not found in the underlying basic sciences, such as the very concepts of planet, atmosphere, and ocean; of plate, continental drift, and even rock; of earthquake, wind, storm, climate, and so on. In short, the basic sciences are necessary but not sufficient for the corresponding applications.

The common acceptation of 'applied science' is the second, and the one we shall deal with, namely the investigation of cognitive problems with *possible* practical relevance. Here is a *pêle-mêle* list of fields of applied scientific research: materials science, in particular metallurgy; the analysis of natural products with a view to isolating some useful chemicals; the experimental synthesis of promising polymers (e.g. synthetic fibers); the investigation of the biology of plants with a possible industrial interest, such as the *guayule* (that produces rubber); the whole of pharmacology and food science; the entire field of medical research, from internal medicine to neurology; psychiatry and clinical psychology; the entire science of education; and the applications of biology, psychology and social science to investigating social problems such as unemployment and marginality, drug addiction and criminality, cultural deprivation and political apathy, with a view to designing social programs aiming at their eradication.

Applied science, in this second sense, does not follow automatically from basic science either. Doing applied science, like doing basic science, is conducting research aiming at acquiring new knowledge. The differences between applied research and basic research are differences in intellectual debt to basic science, in scope, and in aim. The first difference consists in that all applied science in the second sense is also applied science in the first, though not conversely. That is, applied research employs knowledge acquired in basic research. This does not entail that applied research is necessarily a matter of routine: if it did not yield new knowledge it would not be research proper, and to deliver new knowledge it must investigate problems of its own. However, the task of the applied scientist is to enrich and exploit a fund of knowledge that has been produced by basic research.

For example, the space scientists who developed new materials and studied human physiology and psychology under unfamiliar conditions (zero gravity, isolation, etc.) tackled new problems that could not possibly be solved with the sole help of existing knowledge—and they came up with new knowledge. The applied scientist is supposed to make discoveries but is not expected to discover any deep properties or general laws. He does not intend to.

Secondly, the domain or scope of applied science is narrower than that of basic science. For example, instead of studying learning in general, the applied psychologist may study the learning of a given language by a given human group in particular social circumstances—e.g. the way Mexican Americans learn English in a Los Angeles slum. And instead of studying social cohesion in general, the applied sociologists may study the social cohesion of, say, a marginal community in a Latin American shanty town, with an eye on possible ways of improving its lot.

Thirdly, all applied research has some practical target, even if it is a long term one. For example, the forester is not only interested in forests in general but also, and particularly, in forests with a possible industrial utility. And the pharmacologist, unlike the biochemist, focuses his research on chemicals that may be beneficial or noxious to certain species, particularly ours. In every case we expect from the applied scientist to end up every one of his reports by asserting, not just that he has discovered *X*, but that he has discovered *X*, which seems to be useful to produce a useful *Y* or prevent a noxious *Z*. But we will not ask him to design an artifact or a procedure for actually producing *Y* or preventing *Z*: this is a task for the technologist. Table 14.1 illustrates what has been said so far and prepares the ground for what comes next.

Applied science lies between science and technology, but there are no borderlines between the three domains: each shades into the other. The outcome of a piece of basic research may suggest an applied research project, which in turn may point to some technological project. Once in a blue moon a single investigator spans all three. More often than not each task demands people with peculiar backgrounds, interests, and goals. Whereas the original scientist, whether basic or applied, is essentially a discoverer, the original technologist is essentially an inventor of artificial things or processes. Indeed the invention of artifacts or of social organizations is the very hub of technological innovation. No matter how modest, an invention is something new, that did not exist before, or that existed but was beyond human control. Thus fire existed before man, but

TABLE 14.1
Some applied, technological and economic partners of some basic investigations

Basic science	*Applied sciences*	*Technologies*	*Industry, commerce, services*
Mathematics	All	All	Consulting
Astronomy	Optics of light, radio, and X-ray telescopes; photometry, bolometry, etc.	Design of astronomical instruments; invention and development of photographic plates; observatory architecture; etc.	Optical industry; photographic industry; maintenance and repair of astronomic instruments; etc.
Atomic physics	Physics of semiconductors, electronics.	Design of radios, TV sets, computers, calculators, etc.	Manufacture and maintenance of electronic apparata.
Chemistry	Hydrocarbon chemistry	Petroleum engineering (prospection, drilling, refining, etc.).	Building and maintenance of petroleum machinery, installation and maintenance of oil wells and refining plants.
Biology	Biology of edible plants	Phytotechnology of edible plants: creation of new varieties, study of new cultivation methods, etc.	Agriculture, food industry, food marketing.
Sociology	Development sociology	Development plans.	Implementation of development plans.

only man invented ways of producing, maintaining and extinguishing it at will; likewise nuclear energy has always existed, but man tapped it first in 1945. Whereas the invention of fire-making and keeping devices required no science at all, that of the nuclear reactor and the nuclear bomb were technological feats indebted to nuclear physics and the chemistry of fissionable materials, which originated in the disinterested investigations of Becquerel, the Curies, Rutherford, Fermi, and others.

Most of the inventions proposed until the beginning of the Modern period owe hardly anything to science: recall the domestication of most plants, animals, fungi, and bacteria; the plough and metallurgy, architecture and coastal navigation. Things started to change in the 17th century and, particularly, since about 1800. The pendulum clock and the Watt governor are based on mechanics; the electric generator and the electric motor, on electrodynamics; the synthetic products used in industry and medicine, on chemistry; the electronic gadgets, on atomic physics; the supercultivars on genetics, and so on. In short, since about 1800 and on the whole, technological breakthroughs have followed scientific discoveries. The most common pattern is nowadays: *scientific paper—applied science report—technological blueprint*. And, as is well known, the time lag between these stages is decreasing rapidly.

Saying that an invention "is based on" a bit of scientific knowledge does not signify that science suffices to produce technology but that it is employed to some extent in designing the artifact or the plan of action. For example, Joseph Henry designed the first electric motor on the basis of his knowledge of electrodynamics (to which he himself had contributed); and Marconi invented the radio by exploiting Maxwell's theory as well as Hertz's experiments. The modern inventor need not know much science but he cannot ignore it altogether, for what is called "the principle" of a modern invention is some property or law usually discovered by some scientific research. Thus the "principle" of jet propulsion is Newton's 3rd law of motion, and the antihistaminic "principle" is the antigen–antibody reaction discovered by the immunologists. What characterizes the inventor is not so much the breadth and depth of his knowledge as his knack for exploiting what he knows, his wondrous imagination, and his great practical sense.

Invention is only the first stage of the technological process. Next comes the so-called development stage—which is where most inventions sink. The blueprint must be translated into a prototype, or a handful of seeds of a new cultivar, or a few milligrams of a new drug, or a plan for a new social organization. Once these artifacts have been produced it is necessary to put them to the test, to see whether they are effective. In the case of a new drug the tests may take some years and millions of dollars.

The third stage is the design of the production, in the case of artifacts, and of the implementation, in the case of social programs. This may demand the building of an entire pilot plant, which poses new problems, which demand new inventions. (In technology, as in basic science and in life in general, one thing leads to another.) Even if built, the pilot plant may not work satisfactorily for some reason or other; and even if it is technically satisfactory it may prove to be too expensive, or to produce socially undesirable effects, such as unemployment or high pollution. No wonder that most of the R&D budget goes into the development stages.

The rule of thumb is this: for every $10 of the total R&D budget, $1 is devoted to basic research, $2 to applied research, and $7 to technological R&D. One reason for this is that only between one-tenth and one-hundredth of the findings of basic research are applicable, a similar fraction of the results of applied research reaches technology, and again only a like fraction of the technological patents ever reaches the market. Thus the fraction of basic research that has an impact on production or the services lies between one thousandth and one millionth. The short-sighted policy maker takes this low practical productivity as an indication that basic research is not worth being supported. The long-sighted policy maker understands this to be an indication that, if we wish to increase productivity, we must enhance creativity and expedite the now often clogged channels between basic science, applied science, and technology.

If the third and last stage of the technological process has been accomplished successfully, production may begin and, eventually, marketing may be undertaken. These two stages may require new inventions and new tests concerning the production as well as its organization and that of the distribution of the product. Such new inputs are likely to be the more novel and frequent, the more original the invention and the more massive the production line and the marketing network. However, these problems usually lie beyond the horizon of the first inventor, unless he happens to be, like Bell, Edison or Land, a businessman himself. Again, this is largely due to differences in personality traits: the inventor is primarily motivated by curiosity and love of tinkering, not by hopes of fortune.

Table 14.2 describes schematically the process that ends up in the market. In the case of an artifact, such as an electronic calculator, or a procedure such as a medical treatment, all stages are gone over. On the

TABLE 14.2
From lab to market.

Activity	*Carried out by*	*Mainly at*	*Output*
Basic research	Basic scientists	Universities	Data, formulas, graphs,
Applied research	Applied scientists	Universities, industrial labs., state labs., etc.	experimental designs, blueprints of scientific instruments, calculations, methods, etc.
Technological invention	Technologists	Industrial labs.	Designs and models of artifacts or procedures.
Development	Technologists and managers	Industrial labs., workshops, and experimental stations	Prototypes, pilot plants, production plans, social programs.
Production	Workers, foremen, technicians, managers, clerks	Industrial plants, farms, mines, forests, seas	
Marketing	Marketing experts, managers, salesmen, ad men, etc.	Offices, stores, warehouses, etc.	Goods or services.

other hand in the case of a more modest product or service, such as a canned food or the organization schema of a cooperative, the scientific stages are usually skipped: not that scientific knowledge is totally absent, but it is borrowed rather than freshly produced. In other words, whereas some inventions call for new investigations, others can be produced with the help of extant knowledge.

Just as administrators and politicians tend to mistake science for technology, so scientists tend to believe that they are the only innovators: that what others do is low grade work, nearly always routine, that anyone can do. This arrogant belief is false: there is innovation in every one of the stages described in Table 14.2. There is innovation not only when an industry or a service is established for the first time, but also in the course of its maintenance, particularly when new circumstances appear. To be sure, one could try to keep an artificial system, such as a factory or a school, unaltered over the years, but this would be foolish in a society in which everything else changes rather quickly. New needs and competition stimulate inventiveness, and innovation may confer advantages both for competition and cooperation. (Besides, not all innovation is elicited by need: much of it is the product of pure curiosity.)

So far we have been concerned with the basic science-applied science-technology-economy flux, typical of the contemporary period. But there is also a constant reflux in the opposite direction. The laboratory worker uses instruments, materials, drugs and even experimental animals produced in mass and uniformly by industry. Applied science and technology supply basic science, new materials, new instruments, and especially new problems. In short, every one of the components in Table 14.2 acts on all the others, not counting the remaining branches of culture and politics. All of these items compose a system that is characteristic of the modern times, namely the *system of production and circulation of knowledge, artifacts, and services.* See Figure 14.1.

Finally, note that we have included in technology far more than the traditional branches of engineering. In fact we conceive of technology as *the design of things or processes of possible practical value to some individuals or groups with the help of knowledge gained in basic or applied research.* The things and processes in question need not be physical or chemical: they can also be biological or social. And the condition that the design make use of scientific knowledge distinguishes contemporary technology from *technics*, or the traditional arts and crafts.

Our definition of "technology" makes room for all the action-oriented fields of knowledge as long as they employ some scientific knowledge. Hence in our view technology subsumes the following fields (Bunge, 1977d):

> *Physiotechnology*: civil, mechanical, electrical, nuclear, and space engineering.
> *Chemotechnology*: industrial chemistry, chemical engineering.
> *Biotechnology*: pharmacology, medicine, dentistry, agronomy, veterinary medicine, bioengineering, genetic engineering, etc.
> *Psychotechnology*: psychiatry; clinical, industrial, commercial and war psychology; education.
> *Sociotechnology*: management science (or operations research), law, city planning, human engineering, military science.
> *General technology*: linear systems and control theory, information sciences, computer science, artificial intelligence.

In conclusion, basic science, applied science and technology have commonalities as well as differences. All three share essentially the same world view, mathematics, and the scientific method. (For the relatively minor differences among them see Vol. 6.) They differ mainly in their *aims*: that of basic science is to understand the world in terms of patterns; that of applied research is to use this understanding to make further inquiries that may prove practically useful; and that of technology is to control and change reality through the design of artificial systems and plans of action based on scientific knowledge.

3. THE KNOWLEDGE SYSTEM

3.1. *Inter-relations*

Our very definition of a research field (Section. 1.2) includes the condition of a close connection with contiguous fields of inquiry. (On the other hand the pseudo-sciences are typically isolated from the remaining fields of knowledge.) In particular, the various basic and applied sciences and the technologies are closely related to one another because they borrow from each other's funds of knowledge and share a world view, mathematics, and a method of inquiry.

Moreover, although every level of reality—physical, chemical, biological, social, and technical—can and must be studied by itself, the supraphysical levels cannot be adequately understood except with the help of the sciences that study the underlying levels. (See Vol. 4, Ch. 1 for the concept of a level.) The reason is that the lower level sciences study the components of the systems investigated by the higher level ones. Thus physics studies the components of molecules, chemistry those of cells, and physics, chemistry and biology study the components of social systems, namely persons and artifacts. Thus knowledge, and the study of knowledge—i.e. epistemology—must match the level structure of reality. (Hence epistemology presupposes ontology.)

In other words, the composition of a sociosystem is included in the union of the reference classes of the life sciences, chemistry, and physics. (Which is very different from saying that social systems constitute a special class of physical systems—a physicalist thesis rejected in Vol. 4.) Hence the proper study of social systems presupposes some (by no means all) of the lower level sciences. Likewise the composition of a biosystem is included in the union of the reference classes of chemistry and physics. (Again, this is not to say that biosystems are a special class of physical systems—another physicalist thesis rejected in Vol. 4.) Hence a proper study of biosystems presupposes some (not all) chemistry and physics, the sciences that study, *inter alia*, the cell components.

The attentive reader must have noticed that we have stipulated that the study of systems at a given level presupposes *some* lower level sciences, not all of them. For example, the social scientist need not know electromagnetism, for social systems (his objects of study) are neither electrically charged nor magnetized. In other words, the higher level systems do not possess all the properties of their components—but, in compensation, they possess (emergent) properties of their own. Recall Figure 13.3. This is why a higher level science has a specific background properly included in the fund of knowledge of the immediately prior research field (i.e. $B_{n+1} \subset K_n$). The resulting picture of the system of scientific and technological level is a staircase: Figure 14.2. This picture has the disadvantage of being static. Only a flow diagram can represent the various information flows that interconnect all the specialized research fields, turning them into components of a single S&T (science and technology) system.

Our image of the S&T system fits in with our view of the level structure of reality, i.e. our epistemology harmonizes with our ontology. In particular, the two are emergentist, i.e. they recognize the emergent properties that characterize totalities, and the emergent concepts needed to account for them. To put it negatively, ours is not a reductionist image of knowledge because it does not presuppose a physicalist ontology but, rather, an emergentist one. (See Vol. 4 and Bunge (1981a).)

Ordinarily physicalism and emergentism are held in a rather dogmatic fashion. The only way either of them can be supported is by proving that it is possible, or impossible, to reduce every science to physics. This is a herculean task that nobody seems to be willing to undertake. A far simpler strategy is to try and *disprove* the reducibility thesis in the case of a few selected theories. Thus if we succeed in proving that quantum chemistry, though based on quantum mechanics, is not reducible to the latter, or that genetics, though based on chemistry, is not a chapter of the latter, we shall have disproved physicalism and its epistemological concomitant, namely reductionism. This task, sketched in Ch. 13, Section 2.2, will be undertaken in Vol. 6. It will turn out that full inter-theoretic reduction is the exception rather than the rule. This ruins the simplistic thesis of the unity of science via reduction to physics. Still, we do have a unified S & T system characterized by an enormous diversity of components united by a common *Weltanschauung*, a common mathematics and a common method, as well as by strong information flows.

Because the S&T system is a tightly knit one, none of its components can alter in a significant manner without affecting some other components. Only individual items of information, in particular data, can change gradually and locally. Drastic changes in theory or in method can hardly be circumscribed: sooner or later they result in some overall changes. Therefore the history of every component of the S&T system is influenced by that of some other components. This systemicity of the sciences and technologies is in sharp contrast with the nonsystemic character of ordinary knowledge, which can be altered by little steps, i.e. by the accretion or attrition of individual bits of information. Indeed ordinary knowledge is a heterogeneous aggregate of superstitions and crumbs fallen from the high tables of S & T. Therefore it belongs in the dock not in the court of appeal. However, ordinary knowledge will always be attached to specialized knowledge, for it is the starting point of some research, and because its information vehicle, namely ordinary language, is necessarily used in all fields of knowledge. Therefore we must link ordinary knowledge with the S & T system.

Finally we come to the notion of the total system of human knowledge, composed of ordinary knowledge, technics (specialized artisanal knowledge), basic and applied science, technology, the humanities, and parts of the arts and the ideologies. That these fields do combine into a system is

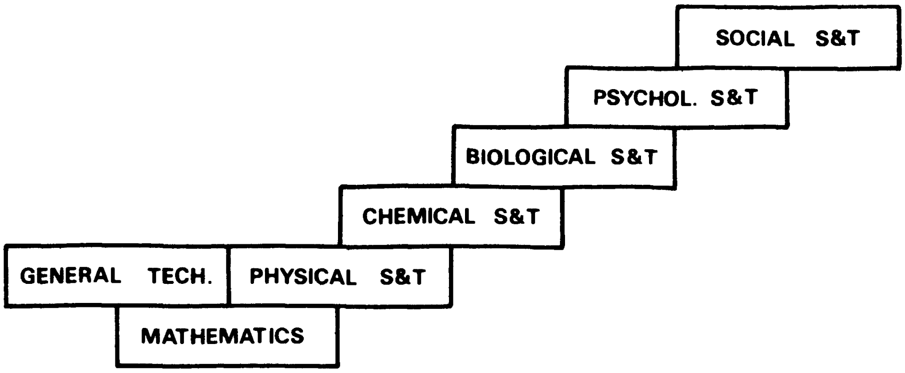


Fig. 14.2. The system of the sciences and technologies. Each step is based on only a part of the preceding one and adds something of its own.

easy to prove. First, ordinary knowledge, despite its limitations, is the *fons et origo* of much specialized knowledge as well as a repository of (often obsolete) stray items of scientific and technological knowledge. Second, science (including mathematics) is the basis of technology as well as a tool and model for the humanities. Third, the technics and technologies serve all and are influenced by all. Fourth, the humanities study, among other things, all kinds of knowledge, and they often draw from the sciences. Fifth, some art genres, in particular the novel and the theatre, give us knowledge of the human condition. Sixth, ideology leads or misleads some research and much technological design. Those seven large fields are then functionally united even though they are logically and methodologically at odds in places. We are therefore justified in speaking of the total *system of human knowledge*: Figure 14.3.

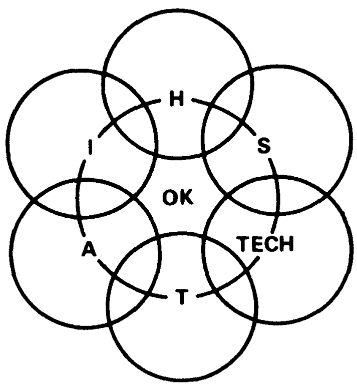


Fig. 14.3. The system of human knowledge: technics (T), technologies (Tech), the sciences (S), the humanities (H), the cognitive part of sociopolitical ideologies (I), and the cognitive part of the arts (A), grow from ordinary knowledge (OK) and enrich one another.

Moreover, as remarked in Section 2.2, knowing is closely related to other human activities, in particular production and the services. Therefore we must introduce the wider notion of the total *system of production and circulation of goods and services of all kinds*, whether cultural, such as theories and books, or noncultural, such as shoes and dental treatments. However, the fact that all these heterogeneous items interact to form a system does not prove that they can be conceptualized in the same manner. In particular, the creation of a new theory cannot be analyzed in the same way as the manufacture of a car.

3.2. *Mergers and Mixtures*

The interrelations among the various components of the human knowledge system enrich and control each of them, and also produce the fusion of some of them. Social psychology is a good example of such fusion: it is not merely the sum of individual psychology and sociology, but a combination of the two. Indeed, it is characterized by constructs and problems of its own that do not occur in either of the parent fields. Two such problems are the changes in personality caused by entry in an unfamiliar social group, and the changes in social organization due to the action of a few individuals. Another interesting synthesis is bioeconomics, which is both a biological and a social science, for it aims at devising methods for the optimal exploitation of renewable resources on the basis of a knowledge of the laws of the latter.

The fusion or merger of research fields is not to be mistaken for their mere addition to form larger fields where each constituent retains its individuality. Such additions, exemplified by space science, geography, and forestry, may be called *mixtures* or *multidisciplines*. (The sum or mixture of two fields of knowledge may be defined as a third field whose first two coordinates equal the physical sums of the corresponding partial coordinates, whereas its eight remaining coordinates are the unions of the corresponding partial coordinates. I.e., $\mathscr{E}_1 + \mathscr{E}_2 = \langle C_1 \dotplus C_2, S_1 \dotplus S_2, D_1 \cup D_2, G_1 \cup G_2, F_1 \cup F_2, B_1 \cup B_2, P_1 \cup P_2, K_1 \cup K_2, A_1 \cup A_2, M_1 \cup M_2 \rangle$, where $\dotplus$ denotes the operation of juxtaposition or physical addition defined in Vol. 3, Ch. 1. If both fields are branches of science, we may set $G_1 \cup G_2 = G_1$, $F_1 \cup F_2 = F_1$, $A_1 \cup A_2 = A_1$, and $M_1 \cup M_2 = M_1$.)

Unlike fusions or mergers, mixtures have no conceptual unity: what unity they do have derives from the referents common to their components fields. Thus space science is a conglomerate of fields united only by their dealing with space ships and their uses. Some of the components of space science are applied sciences whereas others are technologies. Geography is another examples of a multidiscipline, for it is composed of two main fields, physical geography—which is really an earth science—and human geography—which belongs to social science. Because of their great variety, space science, geography, and other multidisciplines—notably medicine—are able to attract workers with a great variety of backgrounds, interests and abilities.

Forestry is another rich multidiscipline encompassing basic sciences, applied sciences, and technologies. A glance at *Forest Science* or the *Journal of Forestry* will attest that the field attracts laboratory and field scientists, mathematical modelers and statisticians, geologists and geographers, bioeconomists and social scientists, and even meteorologists and pychologists. The *basic science* component of forestry is obvious: it is a branch of ecology, for it studies the forest as a community of organisms, and yields knowledge not to be found in the branches of biology concerned with individual organisms. (*Examples*: study of competition among trees, of cooperation among the latter and the undergrowth, and of the coevolution of trees and animals.) The *applied science* component of forestry is no less evident: foresters make use of basic biology, from genetics and cellular biology to plant physiology and ecology, in studying problems of their own that may lead to results of practical interest, such as new ways of intensifying or inhibiting pollination, accelerating or retarding growth, controlling pests of all kinds, etc. All such investigations use basic biology but, by posing new problems, they ensue in new findings, some of which are fed back into basic biology.

Finally, the *technological component* of forestry may be regarded as constituted by forest engineering and forest management. While both deal with the maintenance and exploitation of forests, forest engineering handles the mechanical aspects, such as clear cutting and log transportation, whereas forest management deals essentially with forest maintenance, forestation, harvesting, and non-commercial uses. Forest management is about forests as well as the people who use them, from rangers and loggers to timber businessmen and tourists: it concerns the forest-man system. In sum, forestry may be regarded as a three-tier system embedded in the S&T system: Figure 14.4.

Why should it matter whether forestry—or for that matter the life sciences, the psychological sciences, or the social sciences—is pure or a mixture, a science or a technology? It matters for several reasons. Firstly

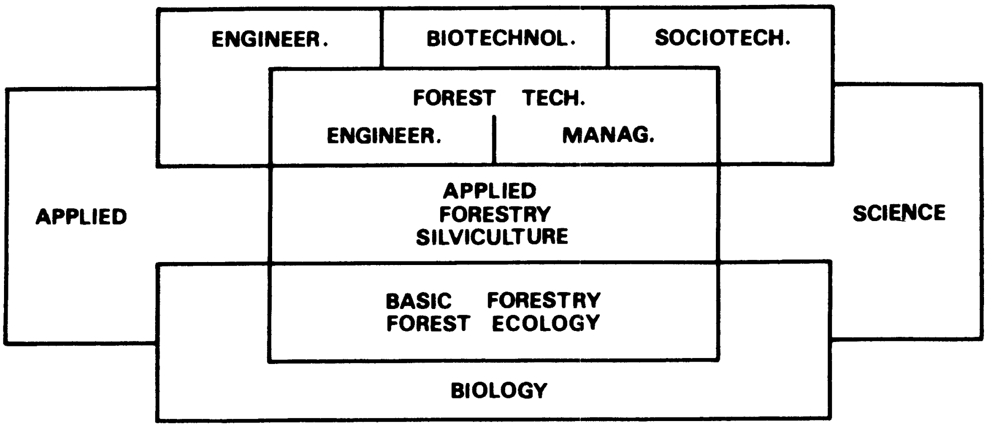


Fig. 14.4. The basic, applied and technological components of forestry.

because, if a field is recognized as a mixture, then it must also be recognized that it is to be cultivated by people with a variety of backgrounds, skills, and interests. (Take note, curriculum designers and science administrators.) Secondly, because a multidiscipline, to stay healthy, must keep open all the communication channels among its components. Thirdly because, if forest management or any other branch of sociotechnology is seen as one component of a rich system with scientific components, it will be realized that any policy and any plan concerning forests—whether it deals with conservation or exploitation—must be based on definite and well tested models of forests rather than on the sole aim of their users. (Thus if the logistic model is used as a rough first approximation, it is realized that the harvesting rate must not exceed the difference between the birth rate and the death rate. Obviously, the criminal harvesting of rain forests has ignored any such model.) In sum, there is no rational action without models and plans based on scientific knowledge. (For a clear pioneering statement of this truism with regard to forest management see Wappes (1926).)

An investigation is said to be *multidisciplinary* if it straddles several research fields, and *interdisciplinary* (or *cross-disciplinary*) if, being multidisciplinary, it does not reach for a collection of separate results but for an integrated view, such as a model interrelating various aspects of the object of study. For example, any reasonable plan for the redesign of a city is the work of an interdisciplinary team composed by geographers, sociologists, economists, city planners, public health experts, teachers, artists, and other specialists exchanging information and views and attacking multidimensional problems.

Most research in science is unidisciplinary or multidisciplinary, seldom interdisciplinary. This is justified in the case of relatively simple objects such as an isolated atom at 0 K. But unidisciplinarity is bound to fail in the case of many-level systems such as chemical reactors, cells, brains, firms, or societies, for they involve several levels of organization. In such cases the prevailing unidisciplinary approach is bound to yield bits of knowledge that "make sense" or "fall into place" only in a wider perspective.

For example, one cannot hope to find out how the brain works with the sole help of neuroanatomy, or neurophysiology, or neuroendocrinology: one needs also evolutionary and developmental studies, as well as individual and social psychology. In practical terms, this entails that the brain should be studied by interdisciplinary teams led by all-rounders. For example, clinical psychologists and neurologists have not learned much about their patients because the anatomy of the latter is usually studied by different workers, namely neuroanatomists and neuropathologists. (Apparently the first cases of Wernicke–Korsakoff syndrome, or severe loss of memory, to have been studied both clinically and anatomically by the same investigators were reported only recently: Mair *et al.*, 1979.) The current fragmentation of fields such as neuroscience, psychology, linguistics, and social science is a major obstacle to their development.

The need for interdisciplinarity is better understood by technologists. The best experts among them are not narrow specialists but multispecialists—not to be mistaken for superficial generalists. That is understandable: the technologist handles things that cannot be artificially isolated from their environment or detached from human action, so he is forced to take many of their aspects into account. He is forced to handle, say, systems of chemical reactions under a variety of conditions, rather than a single reaction under an extreme ideal condition; plants and animals in the field rather than in pots or cages; and people interacting among themselves and solving real life problems, rather than working out academic problems in a psychological laboratory.

The narrow specialist is an endangered species, but nobody should regret its extinction. Things are far too complex to be dealt with by a single discipline. Therefore we need more and more all-rounders of the 17th and 18th centuries savant type. The classical saying about jacks of all trades should be inverted to read *Jack of a single trade—not even master of his own.* After all, the sum total of human knowledge is no mere aggregate: it is a system (Section 2.2). And the systemic character of knowledge matches the systemicity of the world (Vol. 4).

4. Illusory knowledge

4.1. *Pseudoscience*

As defined in Section 2.1, any body of belief and practice that is not scientific but is presented as such deserves being called a *pseudoscience*. There are plenty of pseudosciences, almost as many as genuine sciences. The most popular specimens are parapsychology, psychoanalysis, scientific creationism, Lysenkoist botany, gigo computeering, and monetarism. Pseudoscience is dangerous because (*a*) it passes wild speculation or uncontrolled data for results of scientific research, (*b*) it misrepresents the scientific approach (the "spirit" of science), (*c*) it contaminates some fields of science, particularly the soft and young ones, (*d*) it is accessible to millions of people (whereas genuine science is hard and therefore elitist), and (*e*) it enjoys the support of powerful pressure groups—sometimes entire churches and political parties—and the receptivity of mass media. For all these reasons it behooves philosophers to supply an accurate diagnosis of pseudoscience.

Unfortunately philosophers have so far failed to do so: in some cases they have admitted entire pseudosciences and in others they have rejected whole sciences. This failure is so dismal that one well-known philosopher has concluded that there is no real difference between science and pseudoscience, so that a democratic society should allot equal time to each in all schools (Feyerabend, 1975). The reason for such failure to tell pseudoscience from science is that most philosophers have believed that a single trait—such as empirical content, refutability, or consensus—suffices to characterize science, so that its absence would be sufficient indication of pseudoscience. As we saw in Section 2.1, this is a serious mistake: a field of knowledge must satisfy jointly about ten different conditions to qualify as a science. This is our clue to our definition of a pseudoscience other than the negative definition recalled at the beginning of this subsection.

We stipulate that a field of knowledge

$$\mathscr{E} = \langle C, S, D, G, F, B, P, K, A, M \rangle$$

is a *pseudoscience*, or a *pseudotechnology*, if and only if

(i) C is a community of *believers* who call themselves scientists or technologists although they do not conduct any scientific or technological research;

(ii) The host society S supports C for practical reasons (e.g. because $\mathcal{E}$ is good business) or tolerates C while relegating it beyond the border of its official culture;

(iii) the *domain* D of $\mathcal{E}$ contains unreal or at least not certifiably real entities, such as astral influences, disembodied thoughts, superegos, collective consciousness, national will, UFOs, and the like;

(iv) the *general outlook* G of $\mathcal{E}$ includes either: (*a*) an ontology countenancing immaterial entities or processes, such as disembodied spirits, or (*b*) an epistemology making room for arguments from authority, or for paranormal modes of cognition accessible only to the initiates or to those trained to interpret certain canonical texts, or (*c*) an ethos that, far from facilitating the free search for truth, recommends the staunch defense of dogma, including deception if need be;

(v) the *formal background* F of $\mathcal{E}$ is rather modest. Logic is not always respected, and mathematical modeling is the exception rather than the rule. The few mathematical models that have been proposed (e.g. for psi phenomena) are untestable and therefore phoney;

(vi) the *specific background* B of $\mathcal{E}$ is very small or nil: a pseudoscience learns little or nothing from other fields of knowledge. Likewise it contributes little or nothing to the development of other fields;

(vii) the *problematics* P of $\mathcal{E}$ includes more practical problems concerning human existence (in particular how to feel better and influence other people) than cognitive problems. (Most of the so called pseudosciences are best described as pseudotechnologies.)

(viii) *The fund of knowledge* K of $\mathcal{E}$ is practically stagnant and contains numerous untestable or even false hypotheses in conflict with well confirmed scientific hypotheses. And it contains no universal and well confirmed hypotheses belonging to hypothetico-deductive systems, i.e. law statements;

(ix) the *aims* A of the members of C are often practical rather than cognitive, in consonance with its problematics P. And they do not include the typical goals of scientific research, namely the finding of laws and their use to understand and predict facts;

(x) the *methodics* M of $\mathcal{E}$ contains procedures that are neither checkable by alternative (in particular scientific) procedures nor are they justifiable by well confirmed theories. In particular, criticism is not welcomed by pseudoscientists or pseudotechnologists;

(xi) there is no field of knowledge, except possibly another pseudoscience or pseudotechnology, that overlaps with $\mathcal{E}$ and is thus in a position to enrich and control $\mathcal{E}$. I.e., every pseudoscience and every pseudotechnology is practically *isolated*: there is no such thing as the system of pseudosciences and pseudotechnologies paralleling that of S&T;

(xii) the membership of every one of the last eight components of $\mathcal{E}$ *changes but little* in the course of time and, when it happens to change, it does so in limited respects and as a result of controversy or external pressure rather than research.

We have just described the ideal pseudoscience or pseudotechnology. The description fits psychoanalysis, homeopathy, chiropractice, and other fields. These are all tradition-bound and dogmatic rather than forward-looking and exploratory. (Has anyone ever seen a psychoanalytic laboratory or a mathematical model of the id-ego-superego trinity?) Only parapsychology, which deals with alleged psychic phenomena, is research-oriented. However, 68% of its practitioners have complete, and 22% strong belief in such phenomena (McConnell and Clark, 1980). Besides, parapsychology fails to meet all other conditions for a field of knowledge to be scientific. Let us check them, leaving the details to specialists such as Diaconis (1978), Hansel (1980), and Alcock (1981).

(i) *Domain.* Parapsychology is about immaterial entities, such as disembodied spirits, the existence of which has never been established. So, it is a discipline without a subject matter. On the other hand parapsychology (just as psychoanalysis and mentalistic psychology) ignores the very organ of the mind, namely the brain.

(ii) *General outlook.* The philosopher of science Broad (1949) examined carefully the compatibility of parapsychology with the scientific world view, which he called a "set of basic limiting principles", and concluded that parapsychology does not comply with them—whence they, not parapsychology, had to be given up. For example, precognition violates the principle of antecedence ("causality"), according to which the effect does not happen before the cause. Psychokinesis violates the principle of conservation of energy as well as the postulate that mind cannot act directly on matter. (If it did no experimenter could trust his own readings of his instruments.) Telepathy and precognition are incompatible with the epistemological principle according to which the gaining of factual knowledge requires sense perception at some point.

(iii) *Formal background.* The typical parapsychologist does not excel at handling formal tools, in particular statistics. Thus he consistently selects the evidence ("optional stopping" of a sequence of trials); he does not distinguish a coincidence (accidental or spurious correlation) from a causal relation or a genuine correlation; and he is not fond of mathematical models or even or informal hypothetico-deductive systems: his few hunches are stray.

(iv) *Specific background.* Parapsychology makes no use of any knowledge gained in other fields, such as physics and physiological psychology. Moreover, its hypotheses are inconsistent with some basic assumptions of factual science. In particular, the very idea of a disembodied mental entity is incompatible with physiological psychology; and the claim that signals can be transmitted across space without fading with distance is inconsistent with physics. Worse, parapsychologists brush these inconsistencies aside, claiming that they deal with nonphysical phenomena, so that physicists and other natural scientists are not competent to study them.

(v) *Problematics.* Parapsychology is extremely poor in problems: all its problems boil down to that of establishing that there are paranormal phenomena, i.e. facts that cannot be explained by science. Nor is this problem formulated in clear terms, and this because of the appalling theoretical indigence of parapsychology.

(vi) *Fund of knowledge.* Despite being several thousand years old, and having attracted a large number of researchers over the past hundred years, we owe no single firm finding to parapsychology: no hard data on telepathy, clairvoyance, precognition, or psychokinesis, and no hypotheses to explain these alleged phenomena. All parapsychologists tell us is that these alleged data are anomalous, i.e. unexplained by the science of the day. They suggest no mechanisms and propose no theories. Compare this behavior with that of a scientist, say an astronomer. If an astronomer were to find that a certain celestial object does not seem to "obey" the laws of celestial mechanics or astrophysics, he would feel it his duty to offer or invite some positive conjectures—e.g. that it is not an ordinary body but a quasar or a black hole, a plasma or a laser beam, or some other physical thing. He may conjecture that this thing of a new kind "obeys" laws not yet discovered—but not that it violates well established physical principles such as that of conservation of energy. The parapsychologist does no such thing: he accepts apparently anomalous phenomena as evidence for paranormal abilities, and takes no steps to explain them in terms of laws. Has anyone heard of the First Law of Clairvoyance, or the Second Law of Telepathy, or the Third Law of Psychokinesis? And has anyone ever produced a perpetual motion engine driven by the mind, or a mathematical theory of spooks capable of making definite testable predictions?

(vii) *Aims.* Judging by the accomplishments of parapsychologists, their aim is not that of finding laws and systematizing them into theories in order to understand and forecast. Rather, their aim is either to buttress ancient spiritualist myths or to serve as a surrogate for lost religions.

(viii) *Methodics.* The methods employed by parapsychologists have been scrutinized by scientists, statisticians and stage magicians for more than one century and found almost invariably faulty. The most common defect is lack of strict controls. But deception, either unconscious as in the case of the ordinary experimental subject who wishes the experiment to succeed, or deliberate as in the famous case of the spoon benders, has always plagued parapsychology. (For plenty of amusing examples peruse the journal *The Skeptical Inquirer.*)

(ix) *Systemicity.* Far from being a component of the system of human knowledge, parapsychology is a stray: it does not overlap with any other research field. Therefore its practitioners ask that it be judged on its own merits: on the strength of the empirical evidence they claim to have produced. But this is impossible, quite aside from the fact that such "evidence" is quite suspicious for having been gathered with faulty methods—not to speak of old folk stories and other anecdotal "evidence" still going strong among parapsychologists. Indeed, any fact can be "read" or "interpreted" in a number of ways—i.e. explained by alternative hypotheses. This is one reason that only hypotheses harmonizing with several other hypotheses are worth being investigated. This is not the case with the parapsychological hunches: they do not form a (hypothetico-deductive) system and they do not agree with science. (Recall point (iv).) Moreover, parapsychologists themselves are proud of investigating phenomena, or rather pseudophenomena, that they regard as paranormal for lying beyond the reach of "official" (i.e. ordinary) science.

(x) *Changeability.* Parapsychology cannot be said to be moving fast, the way every genuine science does these days. In fact it is a collection of archaic beliefs that go back to primitive animism. Parapsychologists keep retesting the same hunches over and over again without ever obtaining any conclusive results.

In conclusion, parapsychology is a pseudoscience paragon.

Finally, pseudoscience should not be mistaken for either mere scientific error or unorthodoxy. Science does not have the monopoly of truth: the telephone directories of New York City contain more true statements than all of the social sciences put together—but they happen not to be results of scientific research. While science pursues truth it is not characterized by final truth but rather by—among several other traits—corrigibility. Not so pseudoscience, which is a body of beliefs upheld in the face of either lack of evidence or negative evidence. True, pseudoscientists, like anyone else, accidentally hit on true hypotheses, but they do not investigate them scientifically. Thus Joseph Gall was right in holding that mental functions are discharged by special subsystems of the brain. But there was no experimental evidence for the particular cerebral map he proposed, and therefore no justification for the tenacity with which phrenologists clung to it instead of subjecting it to experimental tests. The same holds for a couple of psychoanalytic hypotheses that may yet turn out to be confirmed by scientific psychology.

Nor should pseudoscience be mistaken for scientific unorthodoxy. In science dissent and controversy are normal and healthy, whereas in pseudoscience they are rare and treated as punishable heresy. Scientific unorthodoxy is just unconventional science; technological unorthodoxy is parallel. Field theory was unorthodox when first proposed because it disagreed with the then dominant action at a distance theories; but it had all the marks of science and it was rife with testable hypotheses and stunning new experiments. Likewise telecommunications, aviation and computer science. All scientific and technological revolutions are, to borrow Isaac Asimov's apt expressions, *endoheresies*—deviations within the S&T system—to be sharply distinguished from *exoheresies* or deviations at variance with S&T. Whether proposed by ordinary members of a research community or by outsiders, a new item of knowledge that fits our definition of science or technology, even though it happens to conflict with some (never all) of the items in the specific background or in the conventional fund of knowledge, qualifies as an endoheresy. Endoheresy, though often resisted, should be welcomed in science and technology, exoheresy should be fought. To paraphrase St. Paul, there is no salvation outside the S&T system.

4.2. *Ideology*

An ideology is a belief system (Ch. 2, Section 3.1) composed of sweeping factual statements and value judgments. We distinguish two genera of ideologies: worldly and unworldly (religious), and two species of the former—world views and sociopolitical ideologies. While everybody holds some *Weltanschauung* (world view) or other, not everybody has religious beliefs or a sociopolitical ideology. Let us examine the various kinds of ideology and the two main epistemological problems they pose, namely their claim to knowledge and their relations to science.

A *world view*, i.e. a body of extremely general beliefs about nature and man, is usually a mixture of deep thoughts and platitudes, well confirmed generalizations and superstitions. In its popular version it often comes in the form of maxims, stories and parables. It is possible to cleanse and systematize any such mixed bag: the result is a philosophy. Whether in the form of a more or less tacit, untidy and inconsistent popular belief system, or in that of an explicit, well-organized and consistent philosophy, a world view is bound to have some cognitive content if only because it concerns the world and our behavior. On the other hand the value judgments and moral norms included in a world view are noncognitive, even though they can be analyzed and even justified in the light of knowledge.

We have argued in Section 2, and also in the Introduction to Vol. 3, that science and technology include certain ontological, epistemological and ethical principles. Hence science and technology may be said to have their own ideology. However, this ideology is not dogmatic: it is being tested and updated by the practice of scientific and technological R&D. Hence it is not ideological in the pejorative sense of the term. (And the rather fashionable claim that science is the ideology of capitalism betrays an utter ignorance of science as well as of the fact that genuine science is roughly the same in all industrialized nations, whether capitalist or socialist.)

A *sociopolitical ideology* is a special kind of world view, namely one concerning only society. And, unlike a general world view, which may be politically neutral, a sociopolitical ideology usually favors the interests of some social group or other, such as a social class or a political party. Beliefs composing a sociopolitical ideology may be grouped into four kinds (Bunge, 1980b): (*a*) *ontological theses* concerning the nature of human individuals, groups, and societies—e.g. answers to such questions as whether persons are spiritual beings, how they get together to form social systems, what makes them conform or rebel, and what are the essential traits of society and its history; (*b*) theses concerning the *economic, political and cultural problems* faced by societies of a given type at a given time; (*c*) *value judgments* about persons and their activities, races, classes, institutions, etc.: in sum, what is good for the individual, the group, or the community; (*d*) *action (or inaction) programs* for the solution (or conservation) of social problems and the attainment (or thwarting) of individual or societal goals.

The first two sets of theses constitute knowledge, sometimes false, but still knowledge. In particular, group (*a*) belongs to some philosophy, whereas group (*b*) is or may become part of social science. On the other hand the value judgments and plans of actions are noncognitive. However, they can, nay ought to, be discussed in the light of scientific knowledge. An example will clarify this point.

Suppose we ask how the Canadian *métis* live and, in particular, what their economic, political and cultural problems are. An a priori answer is likely to be ideological in the pejorative sense of the word. But the question can be investigated scientifically by a team of anthropologists, sociologists and political scientists. Such a team is likely to find that the Canadian *métis* live marginally in poverty and ignorance. Thus, scientific research can confirm or refute ideological assertions concerning social conditions. But if we wonder whether the condition of the Canadian *métis* is fair, we ask for an ideological response—e.g. an inhumane and obscurantist affirmative answer or a human and enlightened negative answer. And no sooner have we admitted the latter than we must ask the further question: Can anything be done about the marginality of the Canadian *métis*, and if so, what? This is an ideological question that poses a political problem.

Now, political problems can be handled ideologically or scientifically. The former procedure consists in adopting some strategy suggested by precedent, intuition, unexamined beliefs, or vested interests. The scientific method consists in designing a social program aimed at solving the problem, in the light of a careful scientific study of the situation. Such study would include not only information about the present condition of the Canadian *métis* but also about the way they would like to live, and the available resources. However, even the most carefully designed social program would fail miserably if it were concocted by well-wishing bureaucrats in Ottawa, out of touch with the *métis*. Successful social programs involve an active participation of the interested party, not only in its implementation but also in its very design. (This again is a result of social research and practice: see Ackoff, 1974.) In sum, some ideological questions can be given scientific answers.

Our example suggests that there are two kinds of sociopolitical ideology: nonscientific and scientific. The former is characteristically dogmatic, incapable of learning from defeat, hence utterly alien to scientific and technological research—ergo anachronistic and utopian. A *scientific ideology*, on the other hand, would be realistic by being inspired by science and by inspiring applied science research as well as sociotechnological design. However, such a possible ideology is still to be built.

A sociopolitical ideology could then join with science to study and solve social problems. But in fact many sociopolitical ideologies have been openly hostile to science or to technology, whereas others have produced grave distortions in the latter. The extreme cases of Nazism *vs.* modern physics, and of Stalinism *vs.* modern biology, are too well known to need comment. A less well known example of the influence of ideology on science is that of the various schools of economic thought: "the mercantilists were the champion of the overseas trader; the physiocrats supported the landlord's interests; Adam Smith and Ricardo put their faith in the capitalist who makes profits in order to reinvest them and expand production. Marx turned their arguments round to defend the workers. Now, Marshall [the greatest of the neoclassical economists] came forward as the champion of the rentier" (Robinson and Eatwell, 1974, p. 39). And, we may add, Milton Friedman pushes monetarism in defense of the near-sighted super-rich and against everyone else.

Nor is ideological influence limited to the social sciences, as the following examples picked at random will show. Spokesmen for the chemical industry defend the hypothesis that cancer is entirely a matter of life style: that it is never caused by the carcinogens released by certain industries and the internal combustion engines. The antipsychiatrists hold that there are no mental disorders but only social ills. Vulgar Marxists claim that every idea has emerged as a product of definite socioeconomic circumstances, and that theories are validated only by social practice. And the partisans of the counter-culture hold that "ethnoscience" (the cognitive part of folklore) is no worse than science. Their favorite argument is the pharmacological knowledge of primitive communities. But a recent study of the Food and Agriculture Organization has disclosed that, out of 20,000 plant species used in folk medicine, only about 200 are effective.

Finally we come to *religion*, i.e. the system of beliefs and practices concerning supernatural agencies and our relations to them. Such beliefs are dogmas, i.e. principles above criticism, and the associated practices are rites of various kinds, mainly rites of worship and purification. Common to all religions is the belief that there are unworldly entities, such as gods and devils—whether purely spiritual or carnal—immortal and endowed with superhuman powers, as well as imperceptible to all but a few chosen individuals, so that their existence must be admitted on faith. The systematic study of such entities is called 'theology', once the very center of intellectual culture. From an epistemological viewpoint the interesting questions with regard to religion and theology are whether they constitute knowledge and whether they are compatible with science.

If knowledge is defined as justifiable true belief, as is commonly done,

TABLE 14.3
Main traits of science and religion, technology and magic.

Component	*Science*	*Religion*	*Technology*	*Magic*
Community	Research community	Church	Tech. community	—
Society	Any modern society	Any society	Any modern society	Any traditional society
Domain	The entire world	Nature and supernature	The manipulable world	Nature and supernature
General outlook	Naturalist ontology, realist epistemology, ethics of free search for truth	Supernaturalist ontology, any epistemology, ethics of acceptance and defence of dogma	Naturalist ontology, opportunistic epistemology, utilitarian ethics	Supernaturalist ontology, instrumentalist epistemology, no ethics.
Formal background	Logic and mathematics	Optional	Logic and mathematics	—
Specific background	As needed	—	As needed	—
Problematics	All cognitive problems concerning the real world	All problems concerning nature and the supernatural	Problems of design and operation	Problems about the control of nature or propiciation of supernature
Fund of knowledge	Growing body of test-factual knowledge, in particular laws	Slowly changing body of dogmas and opinions	Growing body of practical knowledge	Stagnant body of beliefs
Aims	Understanding and predicting with the help of laws	Personal salvation, defense of social status quo	Control with the help of laws	Control
Methodics	Scientific method and scrutable techniques	Revelation, authority, exorcism, incantation, prayer	Scientific method and scrutable techniques	Incantation and exorcism

then theology might qualify as a mode of knowledge—knowledge of the supernatural and of our proper attitude to it. Indeed religious and theological beliefs would be justified (*a*) by revelation to some privileged individuals, such as Moses and Mohammed, and (*b*) by certain canonical texts such as the Bible and the Koran. But we have rejected such definition of "knowledge" (Ch. 2, Section 3.2) because it is possible to know beliefs one does not share, and to believe in items that one does not really know—such as life after death. Moreover we do not admit paranormal modes of cognition (e.g. revelation when in a state of grace) as sources of knowledge, or authority as a valid mode of certifying beliefs. A belief accepted without positive empirical evidence together with the support of the bulk of scientific knowledge is merely a dogma. And dogmatism is not just alien to science, technology and the humanities, but is inimical to them. Therefore religion and theology are methodologically at odds with science, technology and the humanities.

And yet the founders of modern science in the 17th century, almost to a man, from Galilei and Kepler to Boyle and Newton, were sincere albeit somewhat heterodox Christians. Does this not prove that science and religion are compatible? Not at all: it is a mere argument *ad hominem* on the same footing as a commercial exhibiting an athlete endorsing a brand of beer or tobacco. All it shows is that consistency of one's total system of belief is hard to come by, particularly in transition periods and in the midst of a society where the church wields a formidable cultural and political power that can be challenged only at the risk of one's life. The question of the compatibility of religion and science is a matter for methodology, not for history or biography. We wish to know whether the two are compatible *de jure* regardless of the compromises that individuals may work out.

This methodological problem can be solved only by comparing the respective frameworks of religion and science. This is not the place to do so in detail. Table 14.3 must suffice here. It shows that religion is incompatible with science and technology. It is not only that religion is dogmatic rather than critical, and that it accepts revelation but has no use for experiment: if this were all, one might think that the two, though very different, are not mutually contradictory. But they *are* mutually contradictory, for there is no room for supernatural entities in the naturalistic world view that underlies science and technology (Section 2). And this is not a small matter, for it dictates the way some questions are approached and answered. For example, the monotheistic religions tell us that miracles are possible, whereas science and technology recognize only laws and man-made rules;

they hold that the universe was created by God, whereas science admits no creation *ex nihilo*; they maintain that the biospecies have not evolved or, if they have, they have done so under divine guidance rather than by genic changes and natural selection; they maintain that the soul or mind is an immaterial and possibly immortal entity, not a brain function; they claim that "the true religion" originates in God not in man; and, what is worse for the advancement of knowledge, they regard all questions of origins as mysteries not to be tampered with, so that in effect they block research into the most interesting questions.

With the decline of the temporal power of the church in the most advanced countries such theological restrictions on inquiry gradually lost their force to the point that religion, rather than being the supreme judge in intellectual matters, became a subject of scientific research and sometimes even a political defendant. Thus the history, sociology and psychology of religion have found that the supernatural is but a figment of the human imagination, so much so that nearly every society, at least since the Neolithic revolution, has created its own religion or has modified the religious beliefs borrowed from other societies. Furthermore, it has been shown that in many cases the Olympus of a religion is a delayed representation of the social structure (Frankfurt *et al.*, 1946). So, science does have something interesting to say about religion, namely that its myths are in the same epistemological category as those of Aesop's and Disney's fables. The main difference between religious myths on the one hand, and poetic or cinematographic fictions on the other, is social: the former used to discharge some social function (cohesion) or dysfunction (maintenance of privilege).

Humbled by the sensational triumphs of the science of Galilei and Darwin, the authorities of the three monotheistic religions stopped attacking science and tried to reach an accommodation with it. Barring some fundamentalist fanatics, contemporary theologians claim that science and religion are mutually complementary rather than exclusive, and that the condemnations of Galilei and Darwin were merely human errors. According to some, they are compatible because they deal with different questions or "different levels of reality"; to others—in particular Pope John Paul II—because both derive from God. (See, e.g. Barbour (1966, 1968), Bube (1968), Schilling (1973), Schlesinger (1977), MacKay (1978), and the journal *Zygon*.) At least three strategies have been tried to reconcile religion with science: watering down religious dogmas, trying to restrict the reach of science, and distorting the latter.

The first strategy, favored by liberal theologians, consists in abandoning the literal reading of the sacred texts and regarding them as allegories or as what they are, namely fantasies concocted long ago by ignorant men. In particular the book of Genesis, so obviously at variance with scientific knowledge, is brushed aside as a mere fable. But then there is no revealed dogma left, everyone is at liberty to believe what he wishes, and the church is deprived of a firm doctrine as well as of divine authority, as a consequence of which it loses unity and power. Therefore this move is resisted by the fundamentalists, who realize that religion is inimical to free thinking, and who have no patience with the intellectuals who are unable to confess to themselves that they are deists, or perhaps even agnostics, rather than theists.

The second reconciliation strategy consists in holding that "The scientific exploration of the universe, as our late Holy Father [Pius XII] so often emphasized, is good, but it attains its full significance only when it reverently respects God's overlordship. It must be carried out with humility; there must be a real ascesis of knowledge..." (McMullin, 1968, p. 42). In other words, scientists are not to ask any basic questions that may embarrass theologians, and they may not publish any findings contrary to religious dogma. They must be not just modest—i.e. aware of their own ignorance—but humble, i.e. they must undertake only small research projects, leaving the important problems, such as those concerning the evolution of the universe, the origin of life, the nature of mind, the origin of religions, and the social functions of the churches, to the theologians. But of course scientists are naturally curious, so that they do not acknowledge easily any external constraints on their research. This is why they entrench, in the very ethics of scientific research, the right to inquire and question without regard for authority: they are, literally, free thinkers.

The third strategy—accommodating science to religion—is just as unlikely to win the acceptance of scientists, or at least of scientific leaders, for it involves profound distortions of contemporary science. Thus it may involve misreading quantum physics as asserting that it views matter as "manifesting mental, personal and spiritual activities" (Peacocke, 1971, p. 184)—an interpretation of quantum physics that not even the most enthusiastic partisan of the Copenhagen interpretation would subscribe to. It may also involve identifying the beginning of the present phase of the evolution of the universe, i.e. the hypothetical "big bang", with the divine *fiat* (Jaki, 1974; Jastrow, 1978)—a miracle violating all the conservation laws of physics, to say nothing of the naturalistic ontology underlying astronomy. And the accommodation certainly involves claiming that, if biospecies have evolved, then they have done so under divine guidance rather than by themselves—a travesty of evolutionary biology. In turn this entails ignoring that most of the biospecies that there ever were have become extinct—a holocaust difficult to reconcile with the dogma that every creature has been created to some higher purpose by an omniscient and loving Creator.

We conclude that none of the three reconciliation strategies work: that science and religion encroach on one another rather than occupying different territories, and therefore are bound to clash at critical points. Therefore whoever wishes to form a comprehensive and consistent world view must opt for either of them. And whereas those who opt for religion are bound to fight science sooner or later—as St. Paul and St. Augustine did—those who opt for science must be on their guard to prevent religionists from curtailing or distorting science.

To sum up, ideology is at the very core of any culture, so it would be foolish to disregard it and, in particular, to ignore the more or less subtle ways in which it can guide or misguide science and technology. The nonreligious ideologies have a cognitive content in addition to a valuational one, and that content should be rendered explicit and compatible with science and the public interest—or else dropped. On the other hand there is no such thing as religious knowledge: there is only religious belief. But there is of course some knowledge of religion and, in general, of ideology—namely the psychology, sociology, and history of ideologies. So, whereas religion ignores science, science knows religion.

In sum, there is illusory knowledge—in particular pseudoscience and nonscientific ideology—alongside genuine knowledge. When backed by a mass movement or by a government, an ideology can be much more powerful than any pseudoscience or any genuine science. In fact, an ideology can galvanize or paralyze an entire society; and, if obscurantist, it can kill entire branches of science and the humanities. Hence the need for scientists and humanists never to lose sight of ideology, to study it and to keep it in check.

The prescientific ideologies, in particular the great religions, used to provide unified world views where every thing and every event had its place and purpose. By destroying such world views science and secular philosophy have created a vacuum. This vacuum is often filled with an assortment of short-lived cults and pseudosciences. Yet at the same time modern science and philosophy have supplied the building blocks for a new, consistent, largely true and inspiring world view—a world of lawfully changing concrete things that can be known and controlled up to a point.

Modern man is therefore confronted with a trilemma. He may retain a traditional world view and reject science altogether, but at the price of remaining totally alienated from modern culture. Or he may try and fashion a dualistic world view complete with matter alongside deities, laws alongside miracles, souls alongside thinking brains, and blind faith alongside open-eyed inquiry. Or, finally, he may restore the lost unity by adopting the scientific world view, which makes room for the wondrous but not mysterious variety, complexity and variability of the world and our knowledge of it.

5. CONCLUDING REMARKS

There are many fields of knowledge but they can be grouped into ten genera: ordinary knowledge, prescientific technics, pseudoscience, basic science, applied science, technology, the humanities, the sociopolitical ideologies, the arts, and the religions. Each such knowledge genus has its own material and conceptual framework. And each conceptual framework allows one to formulate and investigate certain problems, some of which cannot even be stated in alternative frameworks. (Try to pose a mathematical or physical problem in terms of ordinary knowledge or of religion.)

The various conceptual frameworks have sometimes been regarded as mutually compatible or even complementary. An early version of this view is the medieval doctrine of double truth, philosophical and theological, which proved to be nothing but a trick to gain some freedom of inquiry. A later version was that of James Frazer, of *Golden Bough* fame, according to whom magic, religion and science are continuous for being so many attempts to explain and control the world. A more recent view is the so-called "principle of tolerance" (Carnap, 1950), according to which each framework stipulates its own criteria for the reality of the entities it postulates, so the choice among alternative frameworks is a matter of expediency and fruitfulness rather than truth. An updated version of this view of multiple truths, or rather of nontruth, is epistemological anarchism, according to which science is no better than pseudoscience or religion (Feyerabend, 1975).

To be sure, some conceptual frameworks are mutually compatible with one another. For example, the conceptual frameworks of the plumber and the engineer, of the realist novelist and the sociologist, and of the scientific philosopher and the basic scientist are mutually complementary and even partially overlapping. But others are not. For instance, magic is incompatible with technology, faith healing with medicine, existentialism with logic, psychoanalysis with experimental psychology, and science with ideological or religious dogma. Not only do certain fields compete with others, but some of them are superior to their rivals. For example, magic, religion and pseudoscience are inferior to science and technology as modes of knowledge and guides to action because they do not involve research and do not possess error-correction mechanisms such as analysis and experiment. This is why they have not contributed to shaping the modern world view, which is entirely indebted to science and technology. In short, science and technology are by far the best modes of knowledge and guides to action so far devised by man.

When asserting the superiority of science and technology we do not imply that they are perfect or blameless. Both are and will always remain incomplete (but perfectible); science contains a few pockets of dogma and pseudoscience (but we can spot and eradicate them); and technology is often misguided by immoral goals (as a scientific research into some R & D can reveal). The first point (incompleteness) is obvious, since inquiry is of the very essence of science and technology. If perfection were attainable in these fields, they would become extinct, for they consist in discovering and even making gaps, and in filling them. The second point (dogmatism and pseudoscience) is less obvious. "However alien to science, and not widespread there, dogma *de facto* sometimes infiltrates the realm of research [...] Legitimate disagreement or controversy creates dogma when arguments are no longer listened to" (Öpik, 1977). The tenacity with which some physicists cling to the positivist interpretation of quantum mechanics, and sometimes even attempt to supress dissent, is a case in point. As for the pseudoscientific pockets in science—or rather in the activity of some scientists—suffice it to mention applied catastrophe theory (criticized by Zahler and Sussmann (1977)) and the sociological theories criticized by Andreski (1972). Finally, we all know that not every technological project is undertaken in the public interest. However, unlike ordinary knowledge and the various belief systems, science and technology have the resources for detecting and correcting errors. And the citizens of a genuine democracy have the power to prevent the misuses of technology.

The superiority of the scientific and technological modes of knowledge over all others entails that the former are entitled to pass judgment on alternative modes of knowledge. However, the superiority concerns modes of knowledge, not necessarily bits of knowledge. Thus the farmer may occasionally make more accurate weather forecasts than the meteorologist; the mason may correct the mistakes of the engineer; and the poet may see deeper than the psychologist. But the scientist and the technologist can get to learn more than the farmer, the craftsman or the artist; so, in the end their superior mode of inquiry is bound to yield superior knowledge.

Nor does the superiority of science and technology over the alternative modes of knowledge entail that we can dispense with all of them. We need all seven genera of genuine knowledge: ordinary knowledge and prescientific technics, basic and applied science, the humanities, sociopolitical ideology, and art. But in case of conflict in matters of knowledge the inferior modes of knowledge should defer to the superior ones. In particular, a philosophy marginal to contemporary science and technology, such as Wittgenstein's, or inimical to them, such as Heidegger's, is an anachronism: it has only some historical interest. Philosophy should become increasingly scientific. But we had better leave this subject for the next chapter.

CHAPTER 15

UPSHOT

Knowledge is usually studied in either of four ways: philosophically, psychologically, sociologically, or historically. The philosophical study of knowledge, i.e. epistemology, focuses most of the time on the outcome of inquiry with neglect or even ignorance of that which does the knowing, namely the nervous system of an animal embedded in a society with a definite history.

The psychological study of knowledge, or cognitive psychology, focuses on cognitive abilities, their development and perhaps also their evolution and, in recent times, also their analogies (seldom their disanalogies) with computer information processing. Cognitive psychology is hardly interested in the traditional problems of epistemology and it ignores society as well as history. In the case of the information processing variety of cognitivism, it also ignores the brain.

The sociological study of knowledge, i.e. the sociology of knowledge, investigates the external circumstances of the cognitive process, in particular the nature of the information network and its social matrix. It studies the political and economic circumstances that facilitate or block cognition, and the formal organizations typical of inquiry in modern times. But it is hardly interested in the remaining aspects of the problem of knowledge; it particular it disregards the nervous system and the mechanisms of inquiry, from problem finding through hypothesis framing and data collection to the evaluation of either.

Finally the historical study of knowledge—i.e. the various intellectual histories, such as the histories of philosophy, science, and technology—investigates the evolution of knowledge over the last few thousand years. If deep, it treats the history of knowledge as one aspect of social history. If superficial, it is limited to unveiling particular sequences, such as "*A* taught *B*, who taught *C*, and so on", with disregard for individual motivations and talents as well as for social circumstances.

It should be obvious that every one of these approaches delivers a ghostly product. The first produces an image of knowledge without a knowing subject; the second shows us a knowing subject in a social vacuum and perhaps also with an empty skull; the third depicts a society of brainless individuals pushed and pulled by social forces; and the fourth yields a movie of shadowy individuals or, worse, a catalogue of achievements.

Every one of those partial approaches to the study of knowledge tells us something, though something that makes no sense except in relation to the other three stories. These various partial approaches to the study of knowledge are then mutually complementary rather than exclusive. Indeed the cognitive abilities and activities of the individual knowing subject cannot be adequately understood while ignoring his nervous system and its development, the society he was born into, the nature of the inquiring community he belongs to, and the traditions of the latter.

Therefore only a merger of all four approaches to the inquiry into inquiry—namely the philosophy, psychology, sociology and history of knowledge—can produce an adequate representation of knowledge. This does not entail that there is no place for the philosophical, psychological, sociological, or historical specialist in knowledge: it does entail that none of these specialists should ignore the others or hope to do the entire job by himself. In particular, it is all right to philosophize about knowledge, i.e. to do epistemology, but provided one becomes reasonably familiar with some of the findings of the psychology, sociology, and history of knowledge.

In the preceding chapters we have done some philosophy and psychology of knowledge. In the first section of this chapter we shall have a quick look at the social sciences of knowledge. In the next we shall examine the main arguments pro and con the various epistemological schools. And in the subsequent section we shall gather together a number of maxims, resulting from our preceding work, and constituting a *précis* of our own epistemology and methodology.

1. SOCIAL SCIENCES OF KNOWLEDGE

1.1. *The Descriptive Social Sciences of Knowledge*

Social science can be descriptive ("positive") or prescriptive (normative); in the latter case it is called a policy science. In this section we shall peep at the five descriptive sciences of knowledge: anthropology, sociology, economics, politology, and history. We shall examine knowledge policies in the next section.

The *anthropology of knowledge* studies the peculiarities of communities of inquirers. It is the newest branch of anthropology, that fascinating protoscience. Nearly all of the anthropological studies on communities of inquirers are marred by an extreme externalism, i.e. a total disregard for what makes inquirers tick: the problems they tackle, their motivations and their values. The externalist student usually ignores what his subjects do and is naive enough to believe that he can find it out by studying their behavior. Typically he is a social scientist who studies what goes on in a chemistry (or biology or engineering) laboratory without knowing any chemistry (or biology or engineering). So, his doings are necessarily limited to recording the manipulations of his subjects as well as their small talk. He has no adequate idea of their problems, background knowledge, or methods. Consequently his report is bound to be similar to the deaf music critic's or the blind plastic arts critic's.

The anthropologist intent on studying scientific research or technological design without some background in either is in a far worse situation than when confronted with the rites performed by a primitive tribe. At least he can learn the native language and be able to understand the functions of the ceremonies with the help of his informants, for even the most complicated ceremony is simpler than the simplest scientific theory. Not surprisingly, the invariable result of such externalist anthropology of knowledge is that research and design, far from being the work of interacting individuals, appears as the mysterious result of a collective action. Worse, on occasion facts are said to be constructed by scientists rather than studied by them (Latour and Woolgar, 1979; Knorr, 1982). An Amazonian Indian asked to describe the activities of the anthropologist in residence is likely to do better.

The *sociology of knowledge* investigates the ways society stimulates or inhibits the development of the material and conceptual frameworks of the various fields of knowledge. In particular, it studies the ways social pressure distorts perception, preserves obsolete ideas, and discourages innovation. It investigates the reception—favorable, indifferent, or hostile—of new ideas and practices, in particular the fate of "premature" ideas. It also studies the reciprocal influence of ideas on society—e.g. the impact of science on religious attitudes, and of technology on industry and trade and, through them, on life at large.

The *sociology of knowledge* too has been dominated by externalism, i.e. the view that only social, economic or political circumstances dictate whatever scientists and technologists do or fail to do, irrespective of their peculiar background, competence, and interests. This view, *sociologism*, is not only superficial: it can also be seriously misleading. One good example of the ridiculous extreme to which it has been taken is the thesis, of Platonic and Heglian roots, that individual scientists do nothing but capture and develop the *Zeitgeist*, i.e. the ideas inherent in a given society Thus according to Hessen (1931), Newton's *Principia* were not so much a product of Newton's own brain as "the product of his period". A more recent perversion of the same kind is the claim that the problems of the nature of the mind and of social change cannot even be approached from our present "paradigms", which would be those of the capitalist society. But the original sin was committed by the Marxist authors who generalized their correct statement that some economic theories are bourgeois, to the thesis that the whole of contemporary culture is divided into two irreconcilable camps: bourgeois culture and proletarian culture. Fortunately this dogma seems to have been forgotten. But the main theses of sociologism, namely that entire societies, not individuals, are the knowing subjects, and that the social structure determines the content and validity of knowledge, are still going strong—though supported only by quotations, never by research.

The *economics of knowledge* studies research communities as components of the knowledge industry, and their outputs as commodities. It investigates the half life of technological innovation and its importance as an input to industry and the services. It studies the way nascent industries and services stimulate discovery and invention, and the way vested interests (in particular large corporations) block innovation.

Another central problem in the current discussions among the economists of knowledge is the place of science and technology in society. According to some, knowledge is only a commodity and a factor of production alongside land, energy, capital, labor, and management. (Curiously enough this thesis is shared by the Chicago school, in particular Milton Friedman, and numerous Marxists, who place knowledge in the so-called infrastructure.) This thesis involves a confusion between science and technology as well as a rigid separation between economy and culture. To be sure, some science is applied, and some applied science is an input to technology, which in turn is an input to production—but so are ideology and politics. What is characteristic of inquiry is not that some of it affects the economy but that it may produce *new knowledge* rather than goods or services. In short *some* knowledge—artisanal and technological knowledge—is economically valuable and may therefore be regarded as a merchandise and be priced. But this holds only for a small fraction of the totality of human knowledge: remember that only about 1% of the inventions get patented, and only about 1% of the patents are ever used in production or in the services.

The *politology of knowledge*, a nearly nonexisting discipline, ought to study the relations between the inquiring communities and political power. In particular it ought to study how applied science, technology and the humanities can supply intellectual ammunition for either the preservation or the alteration of the *status quo*; how political power propels or inhibits, guides or misguides scientific, technological and humanistic research; how investigators and educators are encouraged or frightened by political conditions; what are the social responsibilities of the workers in the knowledge industry—e.g. what they can do to enlighten the public and lobby the parliament concerning the arms race, and so on. So far most of the politology of knowledge has been the amateur work of concerned scientists, technologists and educators writing in the *Bulletin of Atomic Scientists* or debating in the Pugwash conferences. The field, tough of paramount importance, is still in its infancy.

On the other hand the *history of science and technology* is the dean of the social sciences of knowledge and therefore so far the most useful of all to the philosopher of knowledge. Still, it is beset by a number of conceptual difficulties. For one thing, until recently it has been frankly partisan—e.g. trying to prove that all research proceeds inductively, or that modern science was born from Protestantism, or from the Industrial Revolution, or even much earlier, from theological disputations in the 14th century. Another characteristic of the field until about 1960 was its extreme internalism, that ignored the larger social and cultural context and even the importance of formal organizations. Over the past few decades the pendulum has swung in the opposite direction, to the point of often ignoring the internal motor of research, i.e. curiosity. And, whether externalists or internalists, most historians of knowledge are still to make up their minds regarding certain key problems. Let us list a few of them.

One such problem is the very origin of modern science and technology: when, where and why did they emerge? Many historians of knowledge do not seem to realize that the answer to this question depends critically upon the definitions of "science" and "technology", which is a philosophical matter. If mere observation, i.e. fact gathering, is meant by "science", then science emerged with the first mammals and birds; if systematic observation handed down by tradition, then the Neolithic peoples or at the latest the Babylonians and the ancient Chinese should get the credit; if the blending of observation and reason in the search for laws, then the Greeks of the classical period; if experiment is added, then we must wait until the beginning of the 17th century. The case of technology is parallel. In any case it makes no sense to rush to answer the question of the origins of X before agreeing on a characterization of X.

A second, allied problem, is who is to count as a scientist or a technologist: anyone who handles scientific or technological problems, anyone who publishes his solutions to such problems, or what? Take for example Boscovich, highly praised by the 1801 edition of the *Encyclopaedia Britannica* and by later scholars such as L. L. Whyte, K. R. Popper and M. Hesse, who have held that Boscovich (*a*) is the immediate forerunner of the later field and atomic theories, and (*b*) his system is an alternative to classical mechanics. Yet a reading of his main work, *Theory of Natural Philosophy* (1758), will persuade any physicist that both claims are mistaken. Indeed, Boscovich's hypothesis that the ultimate constituents of matter are centers of force does not establish a field theory: he did not state any properties of the field, let alone properties of the interactions between fields and bodies. He had neither empirical evidence for his conjecture nor the mathematical tools (partial differential equations) for formulating it. Nor was Boscovich's theory a viable alternative to the Newton-Euler mechanics, for it did not contain any equations of motion, and it postulated a single force law (which, if plugged into Newton's laws, yields no solution in closed form for arbitrary distances). In short, Boscovich's work was not a theory proper, it made no testable predictions, and it was at variance with the best science of his time: it was a clear case of pseudoscience rightly ignored by physicists and textbooks. Why then has it been so highly regarded in our own time by some historians and philosophers of science?

A third problem that often plagues historians of knowledge is whether every discovery or invention is necessary, i.e. unavoidable given the conditions at the time it was made, so that if *X* had not made it then someone else would have done it. Of course such a statement is untestable. We only know that some discoveries and inventions are multiple, but this can be explained by (more or less fortuitous) contemporaries sharing a conceptual framework and struggling with the same problem. The invention of special relativity and the discovery of the double helix structure of DNA are cases in point. In both cases a number of talented people were on the trail, and if the actual winners had not existed someone else *might* (not *would*) have won the race. But this does not apply to the invention of the general theory of relativity or to wave mechanics. These are cases of a one-man race against himself—Einstein in the former case, de Broglie in the latter. In these cases the problem was new, it interested nobody else, and the solution was not only totally unexpected but also fruitful. Moreover society provided no direct input in these cases: it just allowed the investigators to follow their own noses. But the results had eventually a profound and lasting impact on science and, in the case of wave mechanics, even on technology and therefore on industry. It would be extremely artificial to invent a causal chain so rigid that its central links, Einstein and de Broglie, could have been replaced by anyone else.

A fourth problem that still eludes historians of knowledge is the strength of the links between knowledge and society. Take that admirable European 17th century, the cradle of modern science, technology, and philosophy. It was anything but a prosperous and peaceful period: it was a time of economic stagnation and even retreat; it was marked by a number of wars, particularly the savage Thirty Years war, which involved France, Spain, Sweden and particularly Germany, a land which remained totally ruined; and it was a period of witch hunting and burning of heretics on the part of both the Catholic and Protestant churches. Therefore the explanation of the rise of science and technology in that century cannot lie in society, which barely tolerated that splendid knowledge explosion. The explanation must be sought in the history of ideas, beliefs, and attitudes in the immediately preceding period: in the open, adventurous, optimistic approach to inquiry and life generally, that arose in the late Renaissance. Externalism, in particular economic determinism, cannot explain all this.

Externalism of the radical kind, namely sociologism, cannot account for scientific, technological, or humanistic change in general. As we saw in Ch. 13, Section 3, Kuhn (1962), Feyerabend (1975) and others have claimed that different thought styles or paradigms are mutually "incommensurable", to such a point that a person steeped in one of them cannot understand any views belonging to a different style or paradigm. If this were true then the history of knowledge would be impossible, for every historian is presumably immersed in his own "thought community" different from that of the creators he studies. In short, paradigmatism is suicidal.

In conclusion, neither internalism nor externalism does a good job of accounting for the evolution of knowledge. (See Agassi (1981) for criticisms of both.) Nor should this be surprising to a philosopher, for the internalism *vs.* externalism debate is an instance of another two more general debates in philosophy. One is the ontological problem of what determines the behavior of anything: its internal structure, its environment or both? This debate is alive in biology (geneticism *vs.* ecologism), psychology (nature *vs.* nurture), and elsewhere. Obviously the two extremes are mistaken. Everything but the universe as a whole is influenced by something else, but no external action is effective unless it bears on suitable aspects of the thing influenced. However, before a thing can be influenced by external factors it must exist—and it may have come into existence through a process of self-assembly of environmental items. Systemism solves this debate by conceding as much importance to the individual components of a system of inquirers as to their interrelations and their relations to society at large. (See Vol. 4. Ch. 5.)

The internalism-externalism debate is also linked to the traditional controversy between individualists and holists in the philosophy of society. In fact epistemological internalism sides with individualism, for it holds that knowledge is wholly a matter of individuals—empiricists and rationalists nod. On the other hand epistemological externalism comes together with holism, or the doctrine that knowledge somehow resides in society—in agreement with objective idealism and radical sociologism. The *tertium quid* is of course again systemism, or the view that, although only individuals can know and ignore, inquire and doubt, they do so in society, which now stimulates their epistemic activities, now inhibits them.

The internalist approach to the problem of knowledge can achieve only partial successes because no brain can prosper in a social vacuum. And the externalist approach too is partial because not even the most favorable social conditions can make up for the lack of original brains. Only a synthesis of internalism and externalism, of individualism and collectivism, can be hoped to yield a realistic account of knowledge avoiding the extremes of the isolated knowing subject and the brainless community. Such synthesis views cognition as a process occurring in individual brains every one of which is embedded in a society and steeped in a tradition, thus benefiting and suffering from interactions with other inquirers as well as from the suggestions and constraints of past achievements and failures. Such a synthesis can be produced only by a confluence of the philosophy, psychology, and social sciences of knowledge. We must strive then for the constitution of a single field of knowledge of knowledge: an epistemology naturalized, sociologized, and unified, matching the unity of human knowledge we noted in Ch. 14, Section 3.1. See Figure 15.1.

1.2. *Inquiry Policies*

As long as culture was weak and inquiry was conducted by a handful of isolated individuals there was no need to manage it. This is not the case of modern society, where knowledge is pursued vigorously and pervades all sectors of society, to the point that one speaks legitimately of the *knowledge industry*. Like every other industry, the knowledge industry has a management. The knowledge management is usually mixed: it is performed by professional organizations—such as societies of scientists, engi-

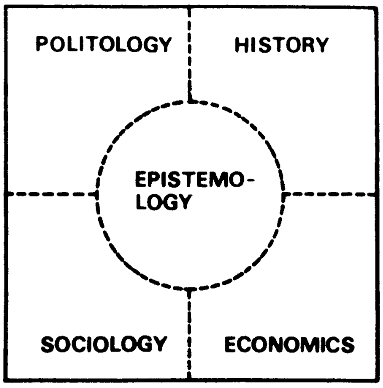


Fig. 15.1. The five aspects of the study of the production, utilization and diffusion of knowledge should be distinguished without being detached. Epistemology is at the center of all the social sciences of knowledge.

neers, or physicians—universities, and whole governmental organizations, not to speak of international bodies such as Unesco.

Like every other management, culture management is supposed to be guided by some policy (or "philosophy" in the vernacular). A policy is a means-ends pair: a set of prescriptions for attaining certain goals with prescribed means, i.e. human and material resources. Like constitutions, policies can be tacit or explicit. A tacit cultural policy is expressed only in practical ways such as the hiring and firing of academic staff in universities and of experts of various kinds in industry and trade; the recruitment (or discouraging) of students; the founding and subsidizing (or throttling) of research projects, museums, libraries, publishing houses, etc. A good unstated but practised cultural policy is better than the best explicit but unpractised policy, which in turn is preferable to an explicit policy aiming at destroying a culture or a sector of it.

Cultural policies, like economic policies, come in two wide genera: noninterventionist ("liberal") and interventionist (planning). The former consist in leaving culture in the hands of private initiative—individual scholars, patrons, foundations, private universities, and the like. In practical terms this means that high grade culture, in particular scientific, technological and humanistic research, is accessible only to wealthy organizations or individuals, for they are the only ones capable of funding such research. This kind of culture freedom—peddled by conservative politicians and economists—is elitist, not democratic. Moreover it curtails freedom of inquiry, for the individual researcher is not supposed to displease his patron. (Remember that Kant's sovereign forced him to take religion seriously.) A democratic society supports its culture on a large scale. And, like any large scale enterprise, it must manage it in a planned way.

Now, planning is double edged: it may stifle or stimulate, according as the planning is authoritarian or democratic. The authoritarian planning of a culture is done from above, i.e. without the active participation of the producers and consumers of cultural goods and services. It is wholly in the hands of bureaucrats or ideologues, who tend to overrate technology at the expense of all the other sectors of culture, and who try to subordinate every cultural activity to the official ideology. Such a planning authority allow creators little or no participation in formulating goals, listing priorities, allocating resources, or even choosing problems and outlining research projects.

Because authoritarian plans are narrow-minded, narrow-sighted, and rigid, they make for mediocrity and inefficiency. This is particularly obvious in the case of scientific, technological, humanistic and artistic creation: the plans for any such activities ought to be outlined by the workers themselves and they ought to allow for sudden changes of technique and even problem. Though obvious, this lesson is still to be learned by many a self-styled democratic government. As Parkinson (1965, p. 116)—of Parkinson's Law fame—notes, "Nowadays, when one country lags scientifically behind another equally prosperous country, the most probable reason is that the government has been telling its scientists what they are to discover. This means, in other words, that too much money has been allocated to specific projects and too little to abstract science. The more resources have been devoted to projects the politician can understand—that is, to the development of discoveries already made and publicized—the fewer resources are available for discoveries which are now inconceivable in so much as they have not yet been made".

A democratic planning of a culture, in particular of its knowledge industry, involves not only freedom of inquiry and teaching but also a balanced stimulation of all the creative sectors of modern culture. Most government officials, particularly if imbued with a narrow ideology, find this difficult to understand: they have no conception of culture, and in particular human knowledge, as a system, and they overrate technology and the services it can render. One must try and teach them that, though technology is indeed central to modern culture, it cannot prosper in isolation from the other branches of culture. One must teach them that advanced creative technology requires basic science; that there is no competitive industry without industrial design, which is an application of art; that there are no effective medical services without medical research, which is applied biology, which rests on basic biology; that social problems cannot be solved without a minimum of sociological research; that there is no science or technology without some philosophy, which in turn interacts with other branches of the humanities; and that there is no cultural creation without an educational system—which must in turn be updated in the light of psychological and sociological research. In short, one must emphasize that the various branches of culture form a system that is intimately linked to the system of production, exchange, and services: that only a systemic vision of culture can help produce an effective cultural policy.

Moreover, given that the clue to success in any social undertaking is active participation (Vol. 4, Ch. 5), a democratic planning of cultural activities will make ample room for such intervention of the producers and consumers of cultural goods and services. Such participation must not be limited to sporadic consultations: the interested parties must belong to the very composition of the culture planning authority. Otherwise the planning will be unrealistic or worse, i.e. stifling.

In short, given the large volume of modern culture, in particular the knowledge industry, culture must be managed on the basis of a plan. A good culture plan (*a*) embraces all the components of the culture system, (*b*) guarantees creative freedom, in particular the freedom of inquiry, (*c*) renders the exercise of such freedom materially possible through generous support of genuine talent, and (*d*) ensures the active participation of the producers and consumers of cultural goods and services in the design of cultural policies and plans as well as in the management of the culture system.

(Whether systemic or fragmentary, democratic or authoritarian, the management of culture, in particular of the knowledge industry, amounts to a control system that regulates the inputs into the culture system. Such control system has a feedback loop that modifies the input rate as a function of the difference between the desired outputs and the actual ones. Assuming that both the inputs or resources and the outputs or cultural goods and services can be reasonably quantitated, the control system can be represented by the following simple-minded mathematical model which neglects, among other things, the time delays, e.g. between problem formulation and problem solution. The model consists of two postulates and their consequences. The first postulate states that the rate $\dot{R}$ of investment of human and material resources is proportional to the difference $D-P$ between the desired value D and the current value P of the cultural production, i.e. $\dot{R}=a(D-P)$. The second postulate is that the rate $\dot{P}$ of production is proportional to the product of P and R, i.e. $\dot{P}=b\,PR$—so that there is neither increase nor decrease in production if there is neither some production to begin with nor some investment. The first factor, a, is a dimensional constant, whereas the second, b, measures the efficiency of the cultural system. The solution to the system of equations is $R^2=(2a/b)\ \ln P^D-(2a/b)P+c$, where c is a constant to be determined by the initial values of R and P. Needless to say, this is only a programmatic model.)

We repeat that cultural planning need not curtail cultural freedom: it will do so only if the plan is authoritarian. Cultural freedom hardly exists when intellectuals and artists depend exclusively on the whim of powerful individual patrons. Nowadays, when culture—in particular knowledge—is mass produced and circulated, often at great expense, cultural planning is indispensable if only to manage its funding. Democratic planning is optimal not only because it can provide the necessary resources to all the components of creative culture, but also because it can protect the individual creators, particularly the more original among them, from the many and assorted foes of free inquiry.

There are two kinds of enemy of free inquiry: external and internal to culture. The former are easily spotted: they are the oppressive bureaucrats, the fanatic ideologists, the pseudoscientists, the military leaders who would like to turn all research into a means of destruction, and the business leaders who would like to restrict inquiry to what promises immediate practical benefits. The internal enemies of free inquiry are less conspicuous but no less obnoxious. They are of three kinds: the researchers who corrupt science to their own personal advantage, the externalist students of knowledge, and the philosophical censors. The former are those who, knowing better, produce low grade, uninteresting and useless results with the sole aim of swelling their *curricula vitarum* and their research grants. Such results, far from improving our understanding or control of reality, produce only an annoying information overload, strain the resources, and make young people bored and disillusioned with the pursuit of knowledge.

The externalist students of knowledge, who are interested only in the social sources and sinks of research, often oppose freedom of inquiry and demand instead that all inquiry have some immediate impact on industry, social services, or government. Obviously only applied science and technology should be assessed by their practical fruits. The main value of basic science, the humanities and the arts is intrinsic not instrumental: they should be judged by their contribution to the enrichment of culture. Therefore, if we wish to save culture we must resist the pragmatist and externalist wind that is sweeping the sciences of science. We must insist that evolutionary biology, cosmology, mathematical logic, and other fields of disinterested research are at least as important as computer science, cancer research and weapons research.

Finally there are the philosophical self-appointed censors of culture. They would like politicians and culture managers to slow down or even suppress all the research that does not conform to their own philosophies. I do not mean just people who disapprove of certain research projects on philosophical grounds, yet take no steps to suppress them. (Classical examples are Comte's disqualification of atomic physics and Bridgman's opposition to field physics.) What I mean is the use of political power, or explicit calls to the powers that be, to interfere in certain research lines or even eliminate them. Two classical examples are Giovanni Gentile's reforms, which all but eliminated the flourishing Italian school of logic, and Zhdanov's attacks on "bourgeois" science and arts. Two more modest and, mercifully, ineffective shows of philosophical censorship of science are Lorenzen's (1967) plea that the state should stop subsidizing abstract mathematics, in particular set theory, and Feyerabend's (1975, p. 216) contention that the social authority of science "has by now become so overpowering *that political interference is necessary to restore a balanced development*" (italics in the original).

2. PHILOSOPHIES OF KNOWLEDGE

2.1. *Rationalism and Empiricism*

By undertaking a philosophical study of knowledge we have tacitly assumed that philosophizing is a mode of knowing: that it may give us some knowledge. (This is far from obvious. Thus irrationalists philosophers do not seek any knowledge, or at least they do not attain any. And Wittgenstein and his followers hold that philosophy is an activity, namely linguistic analysis, to be conceived of as curing the sickness called 'philosophy', so that it yields no knowledge at all—which is true in their case.) And in organizing this *Treatise* we have also assumed that the philosophy of knowledge is among the basic sectors of philosophy: Figure 15.2.

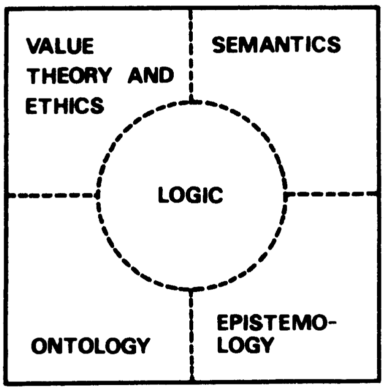


Fig. 15.2. The five sectors of philosophy. Logic is also a part of mathematics and it ought to underlie every branch of philosophy.

As is well known, philosophy, in particular epistemology, far from being homogeneous is divided into a number of schools. Such plurality of schools indicates the hold that tradition has on philosophy, the close links it maintains with ideology, and the rather primitive state of the art. (Controversy is the salt not the meat in matters of knowledge.) In this section we shall review critically some of the living schools of epistemology, note some of their virtues and vices, and place our own theory of knowledge in the map of contemporary epistemology. (For clear discussions of a number of modern theories of knowledge see Blanshard (1939), Hill (1961), and Rorty (1979).)

The two great traditions in epistemology are of course rationalism and empiricism. (Idealism and materialism are ontological schools, each of which can be combined with either rationalism or empiricism.) According to radical rationalism reason, and to radical empiricism experience, is both necessary and sufficient to know what can be known. Rationalism involves the *principle of sufficient reason*, according to which some reason must be given for every statement. Empiricism involves what may be called the *principle of sufficient experience*, on which every statement must be supported by some experience. (Popper and his followers dub *justificationist* anyone who abides by either principle, and they hold that there can be only reasons or experiences against, never for, any ideas. This doctrine, which may be called *negative rationalism*, does not fit the research practice in any field of knowledge: every investigator looks for both positive and negative evidence, conceptual or empirical. Recall Ch. 12.)

Every epistemological school may be regarded as either a special version of rationalism or of empiricism, or as a combination of theses fitting into either of the two great schools. For example, conventionalism is a variety of rationalism, and pragmatism one of empiricism. And kantianism, dialectical materialism, logical empiricism, and our own theory of knowledge are so many syntheses of rationalism and empiricism: Figure 15.3. (Irrationalism does not figure in this diagram because it includes no theory of knowledge.)

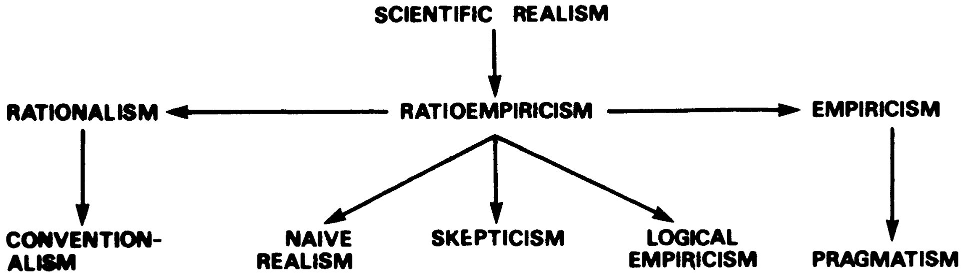


Fig. 15.3. Relations among the main contemporary epistemological schools. The arrow represents the projection operation, whereby a system is reduced to one of its components or one of its versions.

Because of the supremacy it assigns to analysis, theory, proof, and discussion, rationalism is the spontaneous epistemology of most mathematicians and a great many philosophers. Indeed, mathematical and philosophical research are purely conceptual activities—to the extent that these can be dissociated from sensory-motor processes. Rationalism accounts also for the theoretical aspects of scientific, technological and humanistic research. In particular it fits the fact that theory precedes experience—sometimes. On the other hand radical rationalism fails utterly to account for the empirical aspects of inquiry, from perception to experiment: it does not fit the specific activities of the laboratory, field, or workbench inquirer, let alone those of the man of action, for it underrates the value of data and induction; it does not study measurement and experiment—which it regards as, in the best of cases, mere means for weeding out false conjectures; and it exaggerates the importance of deduction and controversy. Finally, the principle of sufficient reason is insufficient: although we ought to give reasons for our hypotheses and even our data, such reasons can be only necessary, not sufficient. More on the inconclusiveness of confirmation and refutation in a while.

Empiricism is the spontaneous philosophy of experimental and field scientists because it extols the virtues of observation, induction, and trial and error. In particular it fits the fact that experience precedes theory—sometimes. Likewise pragmatism—a special kind of empiricism—is the spontaneous epistemology of technicians, technologists, managers and men of action, because of its insistence on the value of practice. On the other hand, because it underrates conceptualization, in particular theorization, empiricism—and particularly pragmatism—is a serious obstacle to theoretical research. And, because of its thesis that every theory is nothing but a summary of data, empiricism also obstructs advanced experimental and technological research, which is guided by theory. Finally, the principle of sufficient experience is insufficient: although we ought to support or undermine our hypotheses with data, these can only be necessary, never sufficient.

We see then that both rationalism and empiricism contain true theses along with false ones. Both reason and experience are necessary yet neither suffices by itself. Kant realized this situation and set out to build a synthesis of rationalism and empiricism. Unfortunately, he combined the bad halves of both. In fact Kant put together the apriorism of rationalism and the phenomenalism of empiricism. (Remember that Kant held that the understanding imposes its laws on nature but that it can know only appearance, not reality. In a way he joined Christian Wolff and David Hume.) Wishing to revolutionize epistemology, Kant effected a genuine counter-revolution.

We have attempted to combine what we take to be the sound halves of the two great epistemological traditions. These are conceptual analysis, theorizing, proof, and discussion, together with observation, measurement, experiment, and praxis—the way Bacon had preached and Galilei practised. The product of this combination may be called *ratioempiricism*. To this we add two other components. One is *critical realism*, which boils down to the thesis that we can build approximately true theories of reality. The other is *scientism*, or the thesis that science is the highest type of knowledge of nature and society, and therefore the best ground for the rational and effective control and enrichment of reality. (More on critical realism and scientism in Section 2.2.) This particular synthesis of rationalism and empiricism, enriched with critical realism and scientism, will be called *scientific realism*. (A few other philosophers call their epistemologies by this name, but they differ from ours in essential respects, particularly with regard to scientism.)

Scientific realism, we claim, accounts for the conceptual components of cognition without being rationalist, and for its empirical components without being empiricist. And it makes room for doubt without falling into skepticism. (We need a strong dose of skepticism to look for errors and be willing to correct them. But we also need to trust the intentions and the veracity of our fellow inquirers if we wish to learn from them. And we must also place some trust, however temporary, in certain guiding principles and methods without which no further research would be possible. Einstein said of Euclidean geometry that it had given the human mind the confidence it needed for its future achievements.)

Scientific realism makes also room for conventions without embracing conventionalism. (Every factual theory involves not only conventions concerning the signification of the symbols it is couched in. It also contains conventions about units and simplifications. Such conventions are necessary to build approximately true representations of real things.) Furthermore, unlike other combinations of rationalism with empiricism—such as kantianism, logical empiricism, and naive realism—our epistemology includes a critical variety of realism. Finally, scientific realism is unabashedly scientistic without being reductionist, in particular physicalist. On the contrary, it acknowledges that every research field has its peculiar concepts and methods, that correspond to the peculiarities (in particular the emergent properties) of its domain or reference class. More in Vol. 6.

Scientific realism is *justificationist* in that it requires every proposition, be it hypothesis or datum, to be ultimately justifiable theoretically or empirically. (A proposition may be said to be *theoretically justified* if it follows from premises in a consistent theory, and *empirically justified* if it is supported by controlled empirical evidence.) However, justification has its limits: it is relative or conditional most of the time. For one thing, the consistency of most theories cannot be proved beyond doubt; and most consistency proofs are relative or conditional, in that they assume the consistency of some other theory. For another, empirical data are just as fallible as hypotheses—but they are equally corrigible and (conditionally) justifiable with the help of some body of (more or less tentative) knowledge. In short, most propositions are not fully justifiable but only conditionally justifiable, and in principle we can always impugn or improve a justification.

Our kind of justificationism is then *fallibilist*, in opposition to the mainstream of epistemology, from Aristotle to Descartes, Husserl, and Dewey, which was always in search of final certainty. Whatever their professed philosophy, scientists and technologists are practising fallibilists, i.e. they believe that theorizing, experimenting, designing and planning are subject to error. But, far from being skeptics, they are *meliorists*: they hope to spot error and reduce it. They do so by rechecking, designing and performing experiments, inviting criticism, etc. So, regardless of how they feel about papal infallibility, scientists and technologists do not believe in their own. (Caution: fallibilism, which surfaced with Peirce, should not be mistaken for falsifiabilism, i.e. Popper's negative rationalism.)

Being fallibilist, our epistemology is not foundationist, i.e. it does not assume that there is a privileged set of unshakeable propositions upon which all the rest can be securely built. (Rationalist foundationism holds that the truths of reason perform that function, whereas empiricist foundationism assigns protocol statements that role, and intuitionism certain allegedly infallible but usually obscure intuitions.) The history of knowledge supplies no evidence for the existence of such a rock bottom. However, this does not entail that there are not even transient foundations. In fact every body of hypotheses (though never of data) can be organized axiomatically, and the set of axioms is rightly called the axiom (or postulate) *basis* (or *foundation*) of that body. What happens is that such basis may not be definitive. Analytical advances may require a reorganization or overhauling, and empirical findings may call for the modification or even rejection of some or all of the axioms. So, *pace* Lakatos (1978, Vol. 2), theories do have foundations, even though these are not final and therefore provide no guarantee of truth. Hence foundations research cannot find certainty but does find order and a greater depth (Hilbert, 1918; Wang, 1966).

Knowledge has no epistemic foundation, whether rational or empirical, but it has a factual or material foundation, namely the real world. Experience is no solid rock but rather shifting sand, and so is theory. In current cognitive practice experience and action mix with theory and analysis, so that neither experience nor reason is supreme or ultimate: each stimulates, supports or corrects the other.

Finally, on scientific realism the traditional subject-object opposition disappears because the knowing (or rather inquiring) subject is seen as being part of the world. Of course one must be able to distinguish facts from ideas about them, but the latter are just further facts, namely those that happen in selected parts of our brains. Scientific realism is thus ontologically monistic in holding that cognition is part of the world—and therefore subject to some of the laws of matter. However, it is pluralistic in that it recognizes a variety of modes of knowledge: ordinary, technical, scientific, technological, and humanistic—though not all of them equally valuable. (Recall Ch. 14.)

(Clearly, the thesis of epistemological monism is ontological, not epistemological. And monism may be idealistic or materialistic, empiricist or neutral: i.e. it may hold that every existent is either ideal, material, experiential, or neither. In line with the ontology expounded in Vols. 3 and 4, we adopt a form of materialist monism. However, this must not be mistaken for semantic materialism, or the thesis that every meaningful statement refers to some chunk of matter. Semantic materialism holds for factual statements but is false of others, particularly the mathematical ones.)

Let us now take a closer look at the two components that distinguish scientific realism from other varieties of ratioempiricism.

2.2. *Realism and Scientism*

By definition factual knowledge is knowledge of facts, i.e. items in the real world, such as states of things and changes in things. The facts that are objects of inquiry can occur inside the inquirer or in his external world, and they can occur spontaneously or as a result of his activity.

Facts are sometimes said to be *hard* in that they do not occur or fail to occur just because we fancy or abhor them. In contrast, our knowledge of facts can be said to be *soft* for being incomplete, at best nearly true, and always corrigible. This distinction between facts in themselves and our images, concepts, or descriptions of them is sometimes blurred in ordinary parlance, as when one speaks of 'getting one's facts straight' when actually meaning "getting an accurate knowledge of the facts that interest one". Occasionally philosophers indulge in this confusion. Thus Feyerabend states that "Facts [probably meaning 'our descriptions of some facts'] contain ideological components" (1975, p. 77), and he rejects the thesis of the autonomy of facts (*ibid.* pp. 38–39).

All scientific and technological inquiry presupposes the autonomy of facts, even of the facts resulting from our own action. (Recall Ch. 4, Section 3.2.) Otherwise, i.e. if we could invent facts, or if we were always unable to distinguish them from fiction, we would not care to check our ideas about them. Error, a mismatch between knowledge and its object, is ever present in cognitive enterprises, and in principle it can always be corrected—if not right now, later on. Knowing that our knowledge of facts may contain errors, and must therefore be checked, presupposes and also confirms the reality or autonomy of the facts in question. (See Bunge (1981), Appendix, for the use of error to criticize subjectivism.). To put it positively: the very search for factual truth presupposes the philosophical theses that there are autonomous facts and that these can be known if only partially. These two theses constitute the nucleus of *realism* or *objectivism.*

Objectivism is the thesis that objectivity is attainable and desirable. Now, objectivity is a property of some perceptions, ideas, or procedures. Actually the word 'objective' designates at least three different concepts, which we shall call 'referential', 'alethic', and 'methodological'. *Referential objectivity* is identical with lack of reference to the knowing subject—as in the statement that there are two people in this room—regardless of its truth value. *Alethic objectivity* is truth for all, as in "I feel pain every time I experience a tooth extraction", which is referentially subjective but alethically objective, for it holds for all the substitution instances of the pronoun. And *methodological objectivity* is reliability together with lack of personal bias in preparing the material of study or performing tests on it, as when using randomization in constructing a statistical sample or conducting studies of performance in evaluating a new artifact or plan.

Natural science, social science and technology are supposed to be referentially, alethically and methodologically objective—i.e. objective *simpliciter*. Objectivity is not easy to attain, particularly when the subject of study is new or one of ideological controversy. Thus it is possible to give a true account of the luminous side of a society while silencing its dark side—a characteristic of propaganda as opposed to social science. A commitment to objectivity is insufficient: we also must strive to account eventually for the *whole* object of our research.

Nor should objectivity be equated with the search for truth. I may truly assert that I am feeling well, and I may claim that all stars are composed of ice. In the first case I state a truth about a subjective state of mind, in the second I am falsely affirming a referentially objective construct. Nor is objectivity identical with the refusal to study subjective experience. Indeed the requirement of objectivity is methodological, not ontological: it entails searching for descriptions the truth value of which does not depend on arbitrary individual fiat, and prescriptions the efficiency of which is likewise independent of fancy. And such research can bear on subjective experience and opinion: witness many psychological and sociological studies of either. (Hayek (1955) confused subjectivism with recognizing the existence and importance of opinions, and consequently declared the social sciences to be subjectivistic.)

Another confusion to be avoided is that between objectivism and impartiality or the refraining from taking sides. If somebody exhibits a certain preference and can somehow justify it in the name of objectivity, then he is being partial without being subjective. For example, anyone engaging in curve fitting prefers order to chaos and hopes that the smooth curve will be truer than the isolated irregular empirical dots. And if a theoretician prefers a deep inaccurate theory to an accurate superficial one, he is not being subjective but partial to deep theories. In either case the scientist is supposed to be able to argue for his preferences instead of taking refuge in dogma. In any case the search for truth involves taking sides for the most promising ideas and methods.

Occasionally scientists forget the requirement of objectivity or even revolt against it. One famous example was von Uexküll's (1928) influential view that "there are as many worlds as there are subjects". He seems to have drawn this conclusion from his own important work on the different ways different animal species perceive their environment. Now, it is true that a scientist and an octopus perceive the world differently, but (*a*) physics, chemistry and biology assume that it is the same world, and (*b*) unlike the octopus, man can become aware of subjective factors and can succeed in overcoming the limitations of perception by constructing objective and even approximately true theories.

Another example of subjectivism is the extreme Copenhagen interpretation of quantum physics, according to which the experimenter "conjures up" all the facts. A related view is Wheeler's (1974) "no-knower-no-world" view, included in his wider "anthropic principle". According to it, the universe was designed and built as man's home—an old religious dogma—so that we are really participants in the foundation and subsistence of the world—presumably a heresy. Of course there is not a shred of evidence for this piece of theological speculation.

The most fashionable vindication of subjectivism rests on a misunderstanding of epistemological realism and a misinterpretation of the recent experimental refutation of Bell's inequalities. Actually these inequalities, which hold for all hidden variables theories, are not so much committed to realism as to two tacit principles of classical physics, which quantum theory does not obey, and which Einstein mistakenly believed to be inherent in epistemological realism. These principles are that (*a*) all physical properties are "sharp" or "well defined" at all times, and (*b*) what happens at a given place can be influenced only by what occurs in its immediate neighborhood (principle of locality or separability). Experiment has refuted the conjunction of these two principles. Realism cannot be refuted experimentally because every well-designed and well-performed experiment involves a clear distinction between object, apparatus, and subject: we must know, at the very least, what is being measured by what and by whom. More in Vol. 7.

In addition to such occasional flares of subjectivism among scientists, the ideal of objectivity has been criticized by two groups of people. The one is formed by the Nazi and Stalinist ideologists who demanded that scientists explode the "myth" of objective research and embrace openly a partisan attitude, i.e. one agreeing with the party line. As is well known, such attacks on objectivism culminated in uncounted personal tragedies and in the banning of entire fields of research. The second group of enemies of the ideal of objectivity is formed by those philosophers, historians and sociologists of science who, following Fleck (1935), Polanyi (1958), Kuhn (1962), and others, claim that all knowledge is personal (meaning subjective), that truth is a social artifact, that the scientist constructs his own object of research, and that the very term "discovery" should be abandoned because it presupposes the autonomous existence of the object of discovery. (See, e.g. Brown, 1977; Barnes, 1982.)

Of course we construct all of our concepts and propositions: these are not found ready made in nature; but we construct models of the world not the world itself. Of course every scientific finding is the end result of a process of inquiry and creation in which imagination is involved; but whatever is discovered—unlike that which is invented—was presumably there to begin with. Even technological inventions are modifications of existents rather than creations out of nothing. Of course original research and invention are subjective processes impossible without passion and hope, fancy and intuition; but in science, technology and the humanities the results of such subjective processes must be checked for truth or usefulness.

If the net results of the "new" epistemology associated with Fleck, Polanyi, Kuhn, Feyerabend, and their followers are that the scientist (or the scientific community) constructs reality instead of modeling it; that there are no objective methods and no objective standards of evaluation; and that society determines what is good or bad science, then this is a *counter-revolution* accompanying the vogue of transcendental meditation, biorhythm, tarot cards, astrology, and parapsychology. It sounds attractive because it emphasizes the subjective aspect of cognition and its social matrix. But it is obscurantist in so far as it rejects objectivity and condones sloppiness. So much for the recent revolt against realism.

Realism is a family of doctrines. Two genera of realism are usually distinguished: ontological and epistemological. *Ontological* (or metaphysical) *realism* asserts the autonomous existence of the world, i.e. its reality independent of the inquirer. (The Platonic variety of ontological realism holds that ideas are the most real of all entities.) *Epistemological realism* maintains that the world can be known. (The scientific variety of realism contains the semantic thesis that scientific theories refer to putatively real entities even though some such entities may turn out not to exist.) Obviously, epistemological realism presupposes ontological realism. Having defended the latter in Vols. 3 and 4, let us turn to epistemological realism.

We distinguish two species of epistemological realism: naive and critical. *Naive* (or *spontaneous*) *realism* holds that the world is just as we see it, i.e. that we know it directly through our senses. On the other hand *critical* (or *constructive*) epistemological *realism* holds that perceptual knowledge is deficient—superficial, incomplete, often wrong; that it must be enriched with conceptual (in particular theoretical) knowledge; and that concepts and their compounds—such as propositions, rules, and theories—are constructions that often go far beyond appearance; and that these constructions represent the world, albeit imperfectly, in a symbolic, not an iconic fashion.

Ordinary knowledge suffices to refute naive realism by noting that one and the same fact may be perceived differently by different subjects. (This argument presupposes that the perceived fact is indeed unique, which of course need not be accepted by the naive realist. But then he runs the risk of switching to subjectivism.) A somewhat more sophisticated argument, used by the later Wittgenstein to attack the naive realism inherent in his *Tractatus*, is that plenty of indispensable concepts, such as those of negation and disjunction, have no real counterparts. (To counter this argument the naive realist may demand that we use a logic deprived of negation and disjunction—which would render discourse extremely poor and clumsy, and argument totally impossible.) A still more sophisticated argument against naive realism is that many important scientific theories refer to unobservable entities and are couched in a complex mathematical language, as a consequence of which they are counterintuitive and cannot be checked by ordinary means. (The naive realist's reply could be that he does not care for such theories.) In conclusion, if we care for science (and technology) we must reject naive realism—not for being realistic but for not being realistic enough, i.e. for not admitting that reality is far more inclusive than appearance.

Realism, then, must be critical or constructive, i.e. it must admit that, in order to know reality, we must invent constructs, in particular theories, that represent reality in devious ways. In fact such theories neither epitomize experience nor represent reality in a point-wise fashion. The first point is clear: theories must be processed and enriched with data before they can represent any empirical facts (e.g. observations). As for the manner in which scientific theories represent reality, we argued in Vol 1, Ch. 3 that it is not the case that every component of a theory has its real counterpart. Rather, the theory *as a whole* represents its referent (or rather some aspects of it) *as a whole*, so the correspondence is global rather than point-wise. For example, the hub of classical electromagnetic theory is the system of equations stating that the d'Alembertian of the four-potential is proportional to the four-density current. In this statement, the d'Alembertian operator is syncategorematic (has no independent meaning); the four-potential is an auxiliary concept (does not represent directly the field); and its analysis into components is (like that of the current) as conventional as the choice of a coordinate system and of a system of units. Only the whole represents (to a first approximation) a system of currents and fields.

Scientific realism is a special version of critical realism, namely the one distinguished by its scientistic component. The word 'scientism' has been assigned several significations (Lalande, 1938). We are interested in two of them. The first is *epistemological scientism*, the thesis that science can give us, and actually often does yield, rather accurate and deep knowledge of reality and, in fact, the best possible factual knowledge. The second is *methodological scientism*, the assertion that the scientific method can and must be tried in all cognitive fields, including the social sciences and humanities.

One may adopt epistemological scientism and not methodological scientism; in fact many natural scientists and many writers on social problems doubt that the latter can be studied in a scientific manner, and some favor "humanistic sociology" and even "humanistic psychology" instead—of which more in Vol. 7. We adopt the two theses of scientism. But we do not accept the reductionist (physicalist) thesis according to which everything biological and social can be reduced to physical and chemical terms: recall Ch. 13, Section 2.2. Nor do we approve of a "slavish imitation of the method and language of Science", which is how von Hayek (1955) defined 'scientism'. Nor, finally, does our variety of scientism require that we accept science in its present state: on the contrary, the attitude of scientism, like that of science, is one of inquiry not belief.

All that scientism states is that scientific research is the best mode of inquiry into any matters of fact—even though it often errs. So, embracing scientism entails accepting alternative modes of inquiry—in particular ordinary knowledge—only provisionally, or because no more accurate or deeper knowledge is needed for a particular purpose. It also entails adopting the philosophical outlook of factual science, which includes not only a realist epistemology but also an ontology of lawfully changing things and the ethos of the free search for truth (Ch. 14, Section 2.1.)

To summarize this section and the preceding one. There are several "theories" (actually views or doctrines) of knowledge, and nearly every one of them contains some truth. In particular, rationalism is adequate for the early phases of rational speculation as well as for formal science, whereas empiricism fits the early phases of factual inquiry. They should therefore be weeded and merged, as well as enriched with principles suggested by the actual practice of inquiry in the most advanced epistemic fields. Two such principles are those of critical (or constructive) realism, and scientism. The resulting synthesis may be called *scientific realism*. It is the epistemology we have tried to build in this book.

We claim that scientific realism is not one more speculative or dogmatic *ism* but the epistemology practised by all scientific investigators in the basic and applied factual sciences, regardless of the philosophy of knowledge they may happen to profess. This we have tried to show in every chapter of this book, by supporting our principles on scientific ways and findings, and criticizing their philosophical rivals on the same grounds. In other words, we presume that our epistemology enjoys a strong inductive support of a very special kind: that of contemporary scientific research. Hence it may fit neither the science of the year 1600 nor that of the year 2200.

For ease of reference we shall now proceed to collect, in the next two sections, some of the epistemological and methodological maxims scattered throughout the previous pages.

3. MAXIMS OF SCIENTIFIC REALISM

3.1. *Descriptive Principles*

The core of any epistemology is composed by a set of descriptive principles every one of which is supposed to epitomize an important aspect of inquiry. Besides having a philosophical value (or disvalue), such principles may contribute to orienting (or disorienting) investigators. Here is a sample of descriptive epistemological maxims gleaned from the previous chapters.

*E*1 The world exists on its own, i.e. whether or not there are any inquirers.

*E*2 We can get to know the world, though only in part, imperfectly, and gradually.

*E*3 Every knowing subject is an animal endowed with a plastic nervous system. (*Corollary*: organisms not so endowed, animal societies, disembodied spirits, machines, etc., cannot know.)

*E*4 Every cognitive act is a process in the nervous system of some animal.

*E*5 Knowing is learning, and the knowledge of an individual is the totality of what it has learned.

*E*6 Learning is the formation of new interneuronal connections (i.e. the constitution of new psychons).

*E*7 Every learning ability develops throughout life and is a result of a long evolutionary process.

*E*8 Humans can know objects of only two kinds: material entities and conceptual objects (constructs).

*E*9 Whereas some constructs derive from percepts, others are invented.

*E*10 An animal can know something about a thing only if the two can be linked by signals that the former can detect.

*E*11 For any given system connectible by signals with an inquirer, the latter can get to know some of the components, some of the things in the environment, and some of the relations in the structure of the system.

*E*12 For any fact physically accessible to an inquirer it is possible to devise means to observe some features of the fact, but there is no means whereby all the traits of any fact can be observed with total accuracy.

*E*13 No inquiry starts from complete ignorance: We must know something before we can formulate a problem and investigate it.

*E*14 Every cognitive operation is subject to error, but every error is corrigible. (*Consequence*: Every correction involves some error, which is in turn corrigible.)

*E*15 There are several ways of knowing: By perceiving, conceiving, and acting; and these various modes combine in many an investigation.

*E*16 All knowledge of factual matters consists in, or involves at some point, some direct observation.

*E*17 Indirect observation yields more and deeper knowledge than direct observation.

*E*18 Some factual knowledge is epitomized in inductive generalizations, and some in hypotheses involving non-observational concepts.

*E*19 Any two humans share some knowledge, but every human knows something that nobody else does. (*Corollary*: Nobody knows everything.)

*E*20 Every animal capable of learning is capable of teaching, e.g. by example.

*E*21 All human inquiry is done in society, and therefore in cooperation and competition with others.

*E*22 All societies set some limits—cultural, political, or economic—on inquiry.

*E*23 All scientific, technological or humanistic inquiry is done nowadays in some research community or other.

*E*24 The main bond that keeps the members of a research community together is the exchange of information, via several languages, with a common goal.

*E*25 Communication among inquirers is the easier, the greater the fund of knowledge, methods and goals they share.

*E*26 In principle every fact and every construct can be investigated: there are problems but no mysteries.

*E*27 Every construct can be clarified. (*Corollary*: Whatever cannot be elucidated is not a conceptual item but merely a noise.)

*E*28 There is no limit to the mathematizability of constructs.

*E*29 Knowledge can be of particulars or of patterns.

*E*30 Knowledge of patterns is compressible into hypotheses and theories.

*E*31 The deeper hypotheses and theories are those involving mechanisms of some sort (not necessarily mechanical ones).

*E*32 In all fields of inquiry we are bound to form causal and probabilistic hypotheses as well as combinations of the two.

*E*33 The hub of every advanced field of factual inquiry is a set of laws (or rather law statements presumably representing objective patterns).

*E*34 Every body of hypotheses can be systematized into a theory, and every theory can be well organized (axiomatized).

*E*35 Every theory proper, when enriched with subsidiary hypotheses and data, can predict, but only mechanismic theories can explain.

*E*36 Every factual theory is a partial representation (global or detailed, true or false to some extent) of supposedly real objects.

*E*37 The testing of a factual theory involves indicator hypotheses relating unobservables to observables.

*E*38 Indicator hypotheses should be checked empirically and justified theoretically.

*E*39 Only theories that can help make predictions are empirically testable.

*E*40 The degree of truth of a theory, and the efficiency of a design, can be found with the help of observation, measurement, experiment, and further theories.

*E*41 Every piece of knowledge and every proposal or design can be improved by research, but not every one is worth being improved on.

*E*42 Advances in knowledge (or in anything else) are sometimes gradual, and at other times they are quick and involve profound alterations.

*E*43 No scientific, technological or humanistic advance comes out of the blue: it always originates in some body of antecedent knowledge.

*E*44 Every novelty in knowledge is comparable with its predecessor, if any, and only such comparison gives us objective grounds for choosing between the two.

*E*45 Every research field is characterized by a general outlook, a background, a problematic, goals and means of its own, as well as by a peculiar research community.

*E*46 Scientific research is the highest of all modes of inquiry.

*E*47 Ordinary knowledge, artisanal knowledge, art, and some socio-political ideologies contain nuggets of genuine knowledge, but they cannot advance any further without the help of scientific, technological or humanistic research.

*E*48 The totality of genuine knowledge constitutes a system, whereas illusory knowledge is marginal.

*E*49 The system of human knowledge is closely connected with the system of production and circulation of goods and services.

*E*50 There are limits, both natural and social, to what man can know, but only the social limits are important—and only these can be overcome.

3.2. *Regulative principles*

Every human inquiry process involves some explicit or tacit epistemological principles. Some of these are regulative, i.e. they guide (or misguide) the planning and execution of inquiry by inspiring problems, methods, hypotheses, or inferences—as well as by suggesting the doubting or rejecting of alternatives, and searching for new principles. Here is a sample of such regulative or methodological principles.

*M*1 Keep studying what others have found, but remember that the quickest and most rewarding way of learning is by conducting independent inquiries.

*M*2 Start your inquiry by choosing a problem that is presumably open, that you feel you can handle, and that may give somebody some satisfaction.

*M*3 If you need a promotion go for the sure thing. If you can afford to be curious, have fun working on tough problems.

*M*4 Do not despise small tasks: Every big problem is surrounded by small problems, and somebody has got to solve them.

*M*5 Formulate your problem clearly: Unearth (or restrict or widen) its context, presuppositions, and data.

*M*6 Do not mistake problems of being for problems of knowledge—e.g. do not try to define causality or free will in terms of predictability.

*M*7 Do not mistake problems of knowledge for problems of being—e.g. do not believe that facts change when seen through alternative conceptual frameworks.

*M*8 Do not let the available techniques dictate all your problems: If necessary try new techniques or even whole new approaches.

*M*9 Plan the investigation into your problem—but be ready to change your plan, and even your problem, as often as necessary.

*M*10 Whenever possible handle your problem scientifically, i.e. armed with scientific knowledge and scientific methods, and aiming at a scientific or technological goal.

*M*11 Do not skirt difficulties—but if you get bogged down switch problems for a while.

*M*12 Do not tolerate obscurity or fuzziness except at the beginning: Try and exactify every key concept or proposition.

*M*13 Do not reify constructs—nor place them in a Platonic realm.

*M*14 Quantitate whatever comes in degrees—but do it only if there is hope of effective measurement, however indirect.

*M*15 Do not commit yourself before checking: First get to know, then believe—and doubt.

*M*16 Revise periodically your most trusted beliefs: You are bound to find fault with some of them.

*M*17 Assign the greatest credence to the best confirmed proposition, the greatest confidence to the most reliable and effective artifact.

*M*18 Regard every principle, method and artifact as fallible in principle—but do not hesitate to use it as long as it does not lead consistently to disaster.

*M*19 Trust your fellow inquirers to pose questions that you did not think of asking, and to find out what you failed to know.

*M*20 Trust your fellow inquirers to make mistakes that you can discover and perhaps correct.

*M*21 Do not try to be totally self-reliant in matters of knowledge: Ask others for information, advice, or help—but feel free not to make use of either.

*M*22 Keep up to date but do not stick to fashion: Every radically new idea, procedure or artifact is unfashionable in the beginning.

*M*23 Listen to your critics but, if you have good grounds to believe that you are on the right track, do not let them bully you.

*M*24 Do not just accumulate data for the sake of increasing the volume of information: Look for patterns.

*M*25 Make liberal use of analogy and induction—but always be on the lookout for their limitations.

*M*26 Simplify—but keep track of your simplifying assumptions and alter them if truth demands it.

*M*27 Make the most of deduction—but check the conclusions to evalute the premises.

*M*28 Do not just accumulate hypotheses: Try to organize them into theories (hypothetico-deductive systems).

*M*29 Prefer theories couched in mathematical terms—but remember that mathematics gives clarity, unity, and deductive power, but not factual truth.

*M*30 A theory a day keeps false data away, and a datum a day keeps false theories away.

*M*31 Watch for errors of all kinds and be ready to correct them if it is worthwhile to do so.

*M*32 Study every entity as a system or a component of such.

*M*33 Recall that every object of study is many-sided and should therefore be approached from several viewpoints.

*M*34 Specialize but never to the point of being unable to understand that an alternative approach is possible.

*M*35 Look for change beneath apparent rest, as well as for the invariants of change.

*M*36 Look for pattern beneath apparent chaos, and for randomness alongside or beneath regularity.

*M*37 Investigate every level in its own right as well as in relation to adjoining levels.

*M*38 Do not skip levels—e.g. do not try to write the Schrödinger equation for the brain.

*M*39 Reduce as far as possible but do not be blind to emergence.

*M*40 Try to integrate all the fields of knowledge that study the same objects.

*M*41 Do not stop trying to explain, but shun theories that presume to explain everything, for they are likely to explain nothing.

*M*42 Do not stop trying to forecast—but do so only with the help of reasonable theories and data.

*M*43 Shun ideology in basic science but watch for it everywhere, and declare frankly your social values in applied science and technology, particularly in social technology.

*M*44 Do not expect to invent any important artifact or social organization without some basic scientific knowledge.

*M*45 Do not expect basic science to deliver technology without further ado.

*M*46 If in a position of scientific or technological power, do not use it except to destroy barriers to discovery and invention.

*M*47 All rules are limited in scope as well as fallible, improvable, or replaceable.

*M*48 Rules of inquiry may help but they do not guarantee success: Original research and design are inventive rather than rule-directed.

*M*49 Don't try to ignore philosophy: Those who ignore philosophy only succeed in reinventing it.

*M*50 In attempting to advance epistemology take into account all the sciences of knowledge.

4. CONCLUDING REMARKS

Our epistemology, scientific realism, is characterized by the following theses formulated and discussed in the preceding pages:

(i) *Realism*: The world exists on its own, i.e. independently of the inquirer, who can succeed in knowing parts and aspects of the world.

(ii) *Naturalism*: Cognition is a brain process, and there are no supernatural or paranormal (i.e. noncerebral) modes of cognition.

(iii) *Evolutionism*: All cognitive abilities evolve.

(iv) *Epistemological socialism*: Human inquiry is done by individuals learning from one another and embedded in society, the norms of which now stimulate, now inhibit research.

(v) *Historicism*: Every inquiry starts from tradition, which it expands and corrects.

(vi) *Moderate rationalism*: Reason is necessary for knowing.

(vii) *Moderate empiricism*: Experience is necessary to know the world.

(viii) *Constructivism*: Concepts and their compounds, even those that represent real things, are our own creation.

(ix) *Conventions*: Some of the concepts occurring in our models of real things are conventional.

(x) *Representation*: Science produces symbolic representations of the world.

(xi) *Praxis*: Technology helps change the world.

(xii) *Justificationism*: Every proposition and every proposal ought to be justified by reason, experience, or in both ways.

(xiii) *Fallibilism*: Every item of factual knowledge, be it datum or hypothesis, is fallible.

(xiv) *Meliorism*: All factual knowledge is perfectible.

(xv) *Scientism*: Anything knowable and worth knowing can be known scientifically or technologically better than in any other way.

(xvi) *Systemism*: The family of research fields, and the body of knowledge reaped in them, each constitute a system.

This concludes our study of general epistemology and methodology. The next volume of our *Treatise* will be devoted to applying those general principles to the investigation of selected topical problems in the philosophy of science and technology.

APPENDIX 3

PARTIAL TRUTH

Our aim in this appendix (revised from Bunge (1981d)) is to find a suitable function from propositions to numbers, such that the value of such function for a given proposition be interpretable as the truth value of that proposition. It is convenient to take the real unit real interval [0, 1] as the range of the function V defined on the set of propositions capable of acquiring a truth value. Thus we shall assume that the truth value function is of the kind $V:P \rightarrow [0, 1]$.

There are many possible choices for V but of course most of them are inadequate. The simplest choice is the usual one, where the range of V is made to collapse into the true-false set. Such theory is inadequate, for it makes no room for half truths. The next simplest choice is to assume $0 \leq V(p) \leq 1$, $V(\neg p) = 1 - V(p)$, and $V(p \vee q) = \frac{1}{2}[V(p) + V(q)]$. This theory is inadequate because it assigns conjunctions the same truth value as disjunctions. (To prove this express the conjunction in terms of disjunction and negation, and make use of the above postulates.)

A third possibility is the theory based on the axioms

$A1$ For all p in P, $V(\neg p) = 1 - V(p)$.

$A2$ For all p and q in P,
(i) $V(p \vee q) = max\{V(p), V(q)\}$,
(ii) $V(p \& q) = min\{V(p), V(q)\}$.

This theory distinguishes correctly between the two basic modes of composition. Besides, it includes a desirable generalization of the *modus ponens*, namely: If p and $p \Rightarrow q$ have the same truth value v, then $V(q) = v$ as well. The trouble with this theory is that a single totally false conjunct suffices to ruin a conjunction, even an infinite one. (*Proof*: Set $V(p) = 0$ in $A2$(ii).) Intuitively, one wishes to say that such a conjunction is partially true rather than totally false.

There are other, more complicated, theories of truth, in particular the ones I proposed in 1963 and 1974a. However, there are problems with them. I have come to suspect that one source of trouble common to all these theories is the way they handle negation. Let me explain. All the theories of truth I am familiar with assume, explicitly or tacitly, that truth and falsity are symmetrical, i.e. that the truth value of not-p is the complement of the truth value of p. This assumption looks intuitive but, like many intuitions, it may turn out to be false. If this is the case then it suffices to ruin all of the extant theories of truth. Let us investigate how the assumption fares in a simple case.

Johnny turns 10 years old. His friend Peter believes Johnny to be 11, whereas his friend Jane suspects that he is 9. Both are in error but not by much: their relative error is 1 in 10, so the truth value of each of their beliefs about Johnny's age can be taken to be $1.0 - 0.1 = 0.9$. Charlie, a third friend of Johnny's, is uncertain about his age and, being a very cautious person, avoids any risky estimates and states "Johnny is *not* 9 years old". He is of course right, and would also be right if Johnny were 8 or 11, 7 or 12, and so on. So, it is a bad mistake to assign his statement the truth value $1.0 - 0.9 = 0.1$, which is what the postulate $\ulcorner V(\neg p) = 1 - V(p) \urcorner$ demands. (In general, when there are more than two possibilities it is less risky to assert a negative proposition. Equivalently, the less informative a proposition, the more chances it has of being true.)

The upshot is that $V(\neg p) \neq 1 - V(p)$. But this negative result of ours is a cheap truth. In order to build a theory of truth we must commit ourselves, i.e. make a positive assumption about the truth value of the negate of a proposition. A simple assumption, to be qualified later on, is this: $\neg p$ is completely false only if p is completely true; otherwise, i.e. if p is less than completely true, $\neg p$ is completely true. In other words, we may assume that, for every proposition p that does have a truth value, $V(\neg p) = 1$, if, and only if, $V(p) < 1$, and 0 otherwise. (We shall add an important qualification later on.) Let us conjoin this assumption with a plausible postulate concerning the composition of propositions, to produce a theory for which we may claim at least partial truth.

We presuppose ordinary logic and stipulate that the truth valuation function $V: P \to [0, 1]$ satisfies the following three axioms.

*A*1 If p is not the negate of another proposition, then

$$V(\neg p) = \begin{cases} 0 & \text{iff} \quad V(p) = 1 \\ 1 & \text{iff} \quad V(p) < 1 \end{cases}$$

Otherwise, i.e. if p is the negate of a proposition q, which in turn is not the negate of another proposition,

$$V(\neg p) = V(q).$$

*A*2 For all p and q in P,

$$V(p\&q) = min\{V(p), V(q)\}$$

*A*3 For all q and p in P,

$$V(p \vee q) = max\{V(p), V(q)\}$$

Let us now extract some consequences and see whether these are plausible.

COROLLARY 1 For all p in P,

$$V(p\&\neg p) = \begin{cases} 0 & \text{iff} \quad V(p) = 0 \quad \text{or} \quad V(p) = 1 \\ V(p) & \text{iff} \quad 0 < V(p) < 1 \end{cases}$$

COROLLARY 2 For all p in P,

$$V(\neg(p\&\neg p)) = 1$$

COROLLARY 3 For all p in P,

$$V(p \vee \neg p) = 1$$

COROLLARY 4 For all p and q in P,

$$V(p \Rightarrow q) = \begin{cases} 1 & \text{iff} \quad V(p) < 1 \\ V(q) & \text{iff} \quad V(p) = 1 \end{cases}$$

THEOREM 1 For all p_i in P, where $1 < i < n$,

$$V\left(\bigwedge_{i=1}^{n} p_i\right) = min\{V(p_1), V(p_2), \ldots, V(p_n)\}.$$

THEOREM 2 For all p_i in P, where $1 \leq i \leq n$,

$$V\left(\bigvee_{i=1}^{n} p_i\right) = max\{V(p_1), V(p_2), \ldots, V(p_n)\}.$$

The first two parts of Axiom 1 seem to capture our considerations on negation, and the third restricts the scope of the standard theorem on double negation. (The latter may warm the heart of intuitionists although it was not meant to.) Axiom 2 and its generalization, Theorem 1, formalize the intuition that the truth value of a conjunction equals the smallest of the truth values of the conjuncts. Axiom 3 and its generalization, Theorem 2, formalize the intuition that disjunction is always safest—if least informative. Corollary 1 says that a contradiction, though absurd, may not be worthless, for it contains at least one worthy proposition. (The trouble with contradiction is not that at best it is half true, but that it entails everything and thus "says" everything that can possibly be said.) Corollary 3 is of course the familiar excluded middle, and thus no part of intuitionism. Corollary 4 generalizes the *modus ponens.*

The theory we have just sketched seems reasonable—to the present writer at the time of writing. It has, however, a defect that may prove to be more than a mere *Schönheitsfehler*, namely that it postulates *A*3 instead of deducing it from the other two axioms. This suggests trying to replace *A*1 or *A*2 with some other assumptions capable of entailing the truth value of a disjunction. And, whether or not this goal is achieved, two further developments are desirable. One is to generalize the theory to encompass the truth valuation of formulas containing quantifiers. (Theorems 1 and 2 may prove useful to this effect.) The other is to build a theory of truth based on the intuitionistic predicate calculus.

The notion of partial truth is used in the next Appendix to elucidate that of predictive power.

APPENDIX 4

PREDICTIVE POWER

Several measures of predictive power are conceivable. The simplest of all is this. If T is a factual theory that is adjoined a set S of subsidiary hypotheses, and is fed a set D of empirical data, then the predictive power of T relative to S and D is the number of different predictions entailed by T together with S and D. A defect of this measure is that it does not involve the accuracy of the predictions. This defect is easily repaired by taking the sum of the truth values of the predictions relative to the new body E of empirical data produced in order to test T and S. Since in general these truth values are comprised between 0 and 1—i.e. the predictions are more or less true rather than either totally true or totally false—we may define the *predictive power* of T relative to S, D and E as

$$\pi(T|SDE) = (1/n)\sum_{1}^{n}{}_{i} V(p_i|E), \text{ where } T \cup S \cup D \vdash p_i,\ 1 \le i \le n, \text{ and } E \cap D = \phi.$$

Clearly, the predictive power of a theory is relative to the additional assumptions and data: a change in either is likely to modify the value of π.

A shortcoming of the above measure is that it makes no difference between new and old predictions, i.e. it does not include the originality or audacity of the theory. The latter can be defined as the dual or complement of the conformity with the relevant antecedent body of knowledge. More precisely, we define the *degree of originality of* proposition p_i relative to both the antecedent body A of knowledge and the fresh empirical evidence E as

$$\omega(p_i|AE) = V(p_i|E) - V(p_i|A),$$

so that the originality is maximal for a true proposition incompatible with the antecedent knowledge, and minimal for a false proposition compliant with the latter. Besides, we rule that the truth value of a proposition be assessed relatively to both the fresh empirical evidence E and the antecedent knowledge A. More precisely we stipulate that

$$V(p_i|AE) = \tfrac{1}{2}[V(p_i|E) + V(p_i|A)],$$

which is comprised between 0 (empirically false heterodoxy) and 1 (empirically true orthodoxy). Finally, we assign each such truth value the weight $\omega(p_i|AE)$, i.e. we multiply truth values by originalities, and add. The result is called the *predictive power* of T relative to S, D, E, and A:

$$\Pi(T|SDEA) = (2/n)\sum_{i=1}^{n} \omega(p_i|AE)\cdot V(p_i|AE)$$

$$= (1/n)\sum_{i=1}^{n} [V^2(p_i|E) - V^2(p_i|A)]$$

The range of this function is $[-1, 1]$. Table A.1 exhibits a few typical values for the a typical case in which all the predictions have the same truth values. True nonconformist theories have the greatest predictive power, whereas false conformist ones have the smallest; true conformist and false nonconformist theories are symmetrical as regards predictability; and a half true heterodoxy has a greater predictive power than a half true orthodoxy.

TABLE A.1

Typical values of the predictive power of a theory T relative to sets S of subsidiary assumptions, D of data used to compute the predictions, E of data used to check them, and A of items of relevant antecedent knowledge.

Theory	$V(p_i\|E)$	$V(p_i\|A)$	$\omega(p_i\|AE)$	$V(p_i\|AE)$	$\Pi(T\|SDEA)$
True conformist	1	1	0	1	0
Half true conformist	$\frac{1}{2}$	1	$-\frac{1}{2}$	$\frac{3}{4}$	$-\frac{3}{4}$
False conformist	0	1	-1	$\frac{1}{2}$	-1
True nonconformist	1	0	1	$\frac{1}{2}$	1
Half true nonconformist	$\frac{1}{2}$	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$
False nonconformist	0	0	0	0	0

APPENDIX 5

FORMAL STRUCTURE OF EXPERIMENT

Let us investigate the algebraic structure of an experiment. We shall consider the physical transactions involved in an experiment, and assume that the conceptual component involved in the design and interpretation is represented in some of those physical processes.

Consider a system composed of an observing thing *a*, or *observer*, and an observed thing *b*, or *object*. The observer system *a* is a person or a team equipped with suitable instruments. The two components can be connected in either of the following ways: *a* acts on *b*, *b* acts on *a*, and *a* and *b* interact. Moreover the couplings can be energetic, informational, or both. If energetic we speak of actions as *inputs*; if informational we speak of *signals*. (But of course every signal is carried by a process that transfers energy.) We are particularly interested in the case where the action of the object on the observer is a signal, i.e. an input to the observer's brain. In this case we have the possible situations shown in Figure A.1. To analyze these three situations we avail ourselves of the concept of state space (Vol. 3, Ch. 3) as well as of a remark of Krohn *et al.* (1967).

Call $S(a)$ and $S(b)$ respectively the state spaces of the observer *a* and the object *b*, Σ a set of signals that *a* is capable of admitting, and E a set of actions or inputs that the object *b* can receive from the subject *a*. Then our three situations are summarized as follows:

Observation (of *b* by *a*)	$\omega : \Sigma \times S(a) \to S(a)$
Action (of *a* on *b*)	$\alpha : E \times S(b) \to S(b)$
Experiment (of *a* on *b*)	$\varepsilon : \langle \alpha, \omega \rangle$

Each observation act is described as follows. When signal $\sigma \in \Sigma$ coming

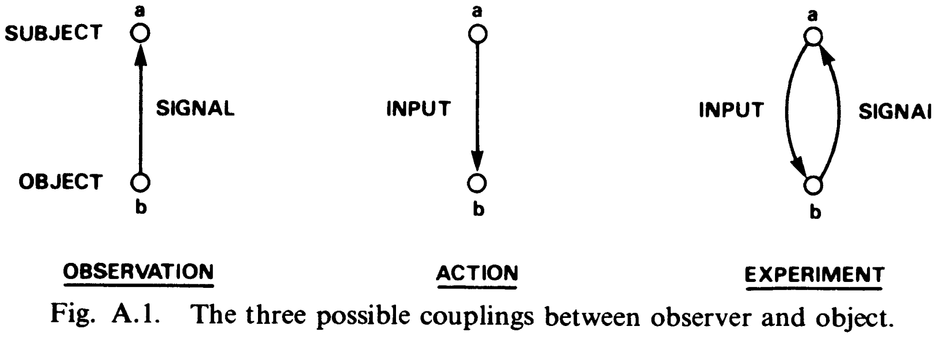


Fig. A.1. The three possible couplings between observer and object.

from object b arrives at observer a in state $s \in S(a)$, a jumps from s to a new state $u = \omega(\sigma, s) \in S(a)$. (We are disregarding time delays, which can easily be incorporated.) Two successive observations combine into one. Thus, if observer a in state $u = \omega(\sigma, s)$ receives signal τ, a jumps to state $v = \omega(\tau, u) = \omega(\tau, \omega(\sigma, s)) \in S(a)$. Hence the set $G(\Sigma) = \{\sigma \in \Sigma | \omega_\sigma\}$ of signal transformations is a semigroup under function composition. Or, if preferred, $\langle S(a), G(\Sigma)\rangle$ is the transformation semigroup acting on $S(a)$. We call it the *observation semigroup*. In other words, the set of all observations performed by a subject on an object with the help of signals of a given kind has the semigroup structure.

What holds for observation holds, *mutatis mutandis*, for action. That is, two successive actions combine to produce a third one. Thus if the object b while in state $f \in S(b)$ receives input $e \in E$ from the observer, it jumps to some other state $g = \alpha(e, f)$. And if b now receives another input h from a, b goes from state g into state $\alpha(h, g) = \alpha(h, \alpha(e, f))$. In terms of the input transformations we have $\alpha_h(g) = \alpha_h(\alpha_e(f)) = \alpha_{he}(f)$. So, actions too form a semigroup, namely $G(E) = \{x \in E | \alpha_x\}$. We call $\langle S(b), G(E)\rangle$ the *action semigroup*.

A while ago we characterized an experiment as a pair of functions, namely an observation function and an action function. We see now that it is more than that, namely an observation semigroup together with an action semigroup. This characterization is correct as far as it goes, but it does not go far enough, for it treats observation and action as mutually independent and on the same footing, which they are not. In fact the inputs to the object do not come from outside the system but are produced by the observer, who is supposed to control them carefully. That is, each input $e \in E$ acting on the object b corresponds to some change $f \in S(a) \times S(a)$ in the state of the observing system or object. (I.e. it is always some event in the observer or its delegate that triggers the experimental action it exerts on the object.) We assume that this correspondence is functional, i.e. that it is represented by a function.

$$k : E \to S(a) \times S(a)$$

to be called the *control function*. The observer event $k(e) \in S(a) \times S(a)$ is called the *controlling event*. Since not all of the changes of state of the observer ensue in controlling events, k is injective (into).

Finally we incorporate a further relation, namely this. Every signal conveys information concerning the state of the object; more precisely, every signal is a message "telling" the observer in a given state what state the object is in as a consequence of the application of an experimental input. We introduce then what we shall call a *messenger function*

$$\mu : \Sigma \times S(a) \rightarrow E \times S(b)$$

such that, for a signal $x \in \Sigma$, and an observer state $s \in S(a)$, $\mu(x,s) = \langle e,t \rangle \in E \times S(b)$ is the message that x conveys to a in state s, namely that the input e is paired off to the object state t.

We collect the various items met so far into the following diagram, where the nameless arrows designate projections.

$$\begin{tikzcd}
S(a) \times S(a) \arrow[r] & S(a) & \Sigma \times S(a) \arrow[l, "\omega"'] \arrow[ddll, "\mu"'] \arrow[dd] \\
E \arrow[u, "\kappa"] & & \\
E \times S(b) \arrow[u] \arrow[rr, "\alpha"'] & & S(b)
\end{tikzcd}$$

So far we have regarded observations and experimental actions as forming semigroups somehow related by the control and messenger functions. Actually they are richer structures, namely categories. In fact, consider the triple $\langle S(a), G(\Sigma), i_a \rangle$, where i_a is the identify morphism (or transformation) in the observer state space $S(a)$. The observer states, i.e. those in $S(a)$, constitute the "objects" of the triple, the observations or signal transformations (i.e. the elements of $G(\Sigma)$) the morphisms, and the null observation (i.e. i_a) the identity morphism of the triple, which is then a category. Likewise in the case of actions: here the "objects" are the object states, the morphisms the experimental inputs, and the null action the identity morphism in the object state space $S(b)$, designated by i_b. In sum we have to do with two categories:

the *category of observations* $\langle S(a), G(\Sigma), i_a \rangle$

and

the *category of actions* $\langle S(b), G(E), i_b \rangle$

which together represent an experiment.

APPENDIX 6

DEGREE OF CONFIRMATION OF A THEORY

To appreciate the magnitude of the inductive leap we take when attributing truth or falsity to a theory imagine an omniscient being confronting a theory T with things of some kind K, such as neutrinos or humans. Being omniscient he knows every one of the infinitely many formulas of T as well as every single possible state of all things of kind K, and he can pair the former off to the latter to find out whether the formulas match the facts. Such omniscient being will find use for the following definitions relating the theory T to a suitable state space S_K for things of kind K:

(i) T *represents* K's (or is a *representation* of things of species K) iff there exists a function ρ: $S_K \rightarrow 2^T$ assigning to every state s in S_K a set t of statements in T;

(ii) T is *true* (or a *true representation* of K's) iff (a) the map ρ is bijective and (b) every member of $\rho(s)$ is totally true for every s in S_K.

Since we, finite beings, do not have access to the totality of formulas of T or to the totality of states of a concrete thing, we must settle for injective maps and partially true propositions in finite numbers. All we can hope for is a high degree of confirmation. This concept can be defined in terms of the concept of partial truth—not probability.

Let T be a specific factual theory and E the totality of empirical data relevant to T such that every member of E is true—a gross simplification. Call E_T the subset of E composed of data joined to T in order to derive testable accounts of facts in a certain domain. The complement $\bar{E}_T$ of E_T is the collection of data employed to check T. When enriched with E_T, T yields n testable consequences $t_1, t_2, \ldots, t_n$. Suppose we check empirically every one of these and ask how well or poorly they account for the available evidence $\bar{E}_T$. If we use the black and white notion of truth we can define the coverage or *degree of confirmation* of T relative to a run of n empirical tests whose outcomes form the set $\bar{E}_T$ as the fraction m/n of hits or confirmations.

A far more realistic measure of the degree of confirmation of T relative to $\bar{E}_T$ is

$$C_n(T|\bar{E}_T) = (1/n) \sum_{i=1}^{n} V(t_i|\bar{E}_T),$$

where $V(t_i|\bar{E}_T)$ is the degree of truth of theoretical consequence t_i judged on the strength of empirical evidence $\bar{E}_T$. Assuming that these degrees of truth range between 0 (total falsity) and 1 (total truth), the values of C_n are similarly comprised between 0 (every prediction is false) and 1 (every prediction is true). An intermediary value such as $\frac{1}{2}$ can be produced in a number of manners, e.g. when every t_i is half true, or half of the t_i's are true and the rest false. The *power* or *volume* of T in accounting for the given facts may be defined as the product of its converage by its depth, or number of levels referred to by T.

BIBLIOGRAPHY

Abraham, R., and J. E. Marsden (1978). *Foundations of Mechanics*, 2nd ed. Reading, Mass.: Benjamin/Cummings.

Ackoff, R. L. (1974). *Redesigning the Future: A Systems Approach to Societal Problems*. New York: Wiley-Interscience.

Ackoff, R. L. (1978). *The Art of Problem Solving*. New York: John Wiley and Sons.

Agassi, J. (1963). *Toward a Historiography of Science. History and Theory, Beiheft 2.* 's-Gravenhage: Mouton & Co.

Agassi, J. (1981). *Science and Society*. Dordrecht and Boston: Reidel.

Alcock, J. (1981). *Parapsychology: Science or Magic*? Oxford: Pergamon Press (cf. 670).

Andreski, S. (1972). *Social Science as Sorcery*. London: André Deutsch.

Bachelard, G. (1938). *La formation de l'esprit scientifique*. Paris: Vrin.

Bachrach, A. J. (1962). *Psychological Research*. New York: Random House.

Balanovski, E., and J. G. Taylor (1978). Can electromagnetism account for extrasensory phenomena? *Nature* **276**: 64–67.

Barber, B. (1962). *Science and the Social Order*. New York: Collier Books.

Barber, B., and W. Hirsch, Eds. (1962). *The Sociology of Science: a Reader*. New York: The Free Press.

Barber, T. X. (1978). Hypnosis, suggestions, and psychosomatic phenomena: a new look from the standpoint of experimental studies. *Am. J. Clinical Hypnosis* **21**: 13–27.

Barber, T. X., and S. C. Wilson (1979). Guided imaging and hypnosis. In A. A. Sheikh and J. T. Shaffer, Eds., *The Potential of Fantasy and Imagination*. New York: Brandon House.

Barbour, I. G. (1966). *Issues in Science and Religion*. Englewood Cliffs, N.J.: Prentice-Hall.

Barbour, I. G., Ed. (1968). *Science and Religion: New Perspectives on the Dialogue*. New York: Harper and Row.

Barnes, B. (1982). *T. S. Kuhn and Social Science*. New York: Columbia University Press.

Bechtereva, N. P. (1978). *The Neurophysiological Aspects of Human Mental Activity*. New York: Oxford University Press.

Békesy, G. von (1967). *Sensory Inhibition*. Princeton: Princeton University Press.

Benado, M., M. Aguilera, O. A. Reig, and F. J. Ayala (1979). Biochemical genetics of chromosome forms of Venezuelan spiny rats. *Genetica* **50,2**: 89–97.

Berk, R. A., D. Rauma, S. L. Messinger, and T. F. Cooley (1981). A test of the stability of punishment hypothesis. *Amer. Sociol. Rev.* **46**: 805–829.

Bindra, D. (1976). *A Theory of Intelligent Behavior*. New York: Wiley Interscience.

Bindra, D., Ed. (1980). *The Brain's Mind: A Neuroscience Perspective on the Mind–Body Problem*. New York: Gardner Press.

Bitterman, M. E. (1975). The comparative analysis of learning. *Science* **188**: 699–709.

Blanshard, B. (1939). *The Nature of Thought*, 2 vols. London: Allen and Unwin.

Bohr, N. (1934). *Atomic Theory and the Description of Nature*. Cambridge: Cambridge University Press.

Bolzano, B. (1837). *Wissenschaftslehre*, 4 vols. Repr.: Lepzig, Meiner, 1929.

Boudon, R. (1967). *L'analyse mathématique des faits sociaux.* Paris: Librairie Plon.

Bourke, J. (1978). *Connections.* Boston: Little, Brown & Co.

Bridgman, P. W. (1927). *The Logic of Modern Physics.* New York: Macmillan.

Broad, C. D. (1949). The relevance of psychical research to philosophy. *Philosophy* **24**: 291–309.

Brown, H. I. (1977). *Perception, Theory and Commitment. The New Philosophy of Science* Chicago: Precedent Publ. Inc.

Bube, R. H. (1968). *The Encounter Between Christianity and Science.* Grand Rapids, Mich.: W. B. Erdman's Publ. Co.

Bunge, M. (1959a). *Causality.* Cambridge, Mass.: Harvard University Press. Rev. ed.: *Causality in Modern Science.* New York: Dover, 1979.

Bunge, M. (1959b). *Metascientific Queries.* Evanston, Ill.: Charles C. Thomas.

Bunge, M. (1962b). An analysis of value. *Mathematicae Notae* **18**: 95–108.

Bunge, M. (1963). *The Myth of Simplicity.* Englewood Cliffs, N.J.: Prentice-Hall.

Bunge, M. (1967a). *Scientific Research I. The Search for System.* New York: Springer-Verlag.

Bunge, M. (1967b). *Scientific Research II. The Search for Truth.* New York: Springer-Verlag.

Bunge, M. (1967c). *Foundations of Physics.* New York: Springer-Verlag.

Bunge, M. (1973a). *Philosophy of Physics.* Dordrecht: Reidel.

Bunge, M. (1973b). *Method, Model and Matter.* Dordrecht: Reidel.

Bunge, M. (1973c). On confusing 'measure' with 'measurement' in the methodology of behavioral science. In M. Bunge, Ed., *The Methodological Unity of Science*, pp. 105–122. Dordrecht: D. Reidel.

Bunge, M. (1974a). *Sense and Reference* (Treatise, Vol. 1). Dordrecht: Reidel.

Bunge, M. (1974b). *Interpretation and Truth* (Treatise, Vol. 2). Dordrecht: Reidel.

Bunge, M. (1976). A model for processes combining competition with cooperation. *Appl. Math. Modell.* **1**: 21–23.

Bunge, M. (1977a). *The Furniture of the World* (Treatise, Vol. 3). Dordrecht and Boston: Reidel.

Bunge, M. (1977b). Levels and reduction. *Am. J. Physiol.* **233**: R75–82.

Bunge, M. (1977c). The interpretation of Heisenberg's inequalities. In H. Pfeiffer, Ed., *Denken und Umdenken: zu Werk und Wirkung von Werner Heisenberg*, pp. 146–156. München-Zürich: Piper & Co.

Bunge, M. (1977d). The philosophical richness of technology. In F. Suppe and P. D. Asquith, Eds., *PSA 1976*, Vol. 2, pp. 153–172. East Lansing, Mich.: Philosophy of Science Assn.

Bunge, M. (1977e). Towards a technoethics. *The Monist* **60**: 96–107.

Bunge, M. (1978a). The limits of science. *Epistemologia* **1**: 11–32. Repr. in *The Physiologist* **23**: 7–13 (1980).

Bunge, M. (1980a). *The Mind–Body Problem.* Oxford and New York: Pergamon Press.

Bunge, M. (1980b). *Ciencia y desarrollo.* Buenos Aires: Siglo Veinte.

Bunge, M. (1981a). *Scientific Materialism.* Dordrecht and Boston: Reidel.

Bunge, M. (1981c). Four concepts of probability. *Appl. Math. Modell.* **5**: 306–312.

Bunge, M. (1981d). Half truths. In E. Morscher and G. Zecha, Eds., *Philosophie als Wissenschaft/Essays in Scientific Philosophy*, pp. 87–91. Bad Reichenhall: Comes Verlag.

Bunge, M. (1981e). Review of Fleck (1935). *Behav. Sci.* **26**: 178–180.

Bunge, M. (1982a). The revival of causality. In G. Fløistad, Ed., *Contemporary Philosophy*, Vol. 2, pp. 133–155. The Hague: Martinus Nijhoff.

Bunge, M. (1982b). *Economía y filosofía.* Madrid: Tecnos.

Bunge, M. (1983). Speculation: wild and sound. *New Ideas in Psychol.* **1**: 3–6.

Carnap, R. (1950). Empiricism, semantics, and ontology. *Revue internationale de philosophie.* **4**: 20–40.

Chomsky, N. (1975). *Reflections on Language*. New York: Pantheon Books.

Claparède, E. (1934). La genèse de l'hypothèse. Étude experimentale. *Archives de psychologie* **24**: 1–155.

Cole, S., J. R. Cole and G. A. S. Simon (1981). Chance and consensus in peer review. *Science* **214**: 881–886.

Coles, J. (1973). *Archaeology by Experiment*. New York: Charles Scribner's Sons.

Coren, S., and J. S. Girgus (1978). *Seeing is Deceiving*. Hillsdale, N.J.: Erlbaum Assoc.

Crick, F. (1966). *Of Molecules and Men*. Seattle: University of Washington Press.

D'Abro, A. (1939). *The Decline of Mechanism (in Modern Physics)*. New York: D. Van Nostrand Co.

Dawes, M. (1980). You can't systematize human judgment: dyslexia. *New Directions for Methodology of Social and Behavioral Sci.* No. 4: 67–78.

Diaconis, P. (1978). Statistical problems in ESP research. *Science* **201**: 131–136.

Eccles, J. C. (1978). *The Human Mystery*. New York: Springer-Verlag.

Eccles, J. C. (1980). *The Human Psyche*. Berlin, Heidelberg, New York: Springer-Verlag.

Einstein, A. (1936). Physics and reality. *J. Franklin Institute* **221**: 313–347.

Falk, R. (1982). On Coincidences. *The Skeptical Inquirer* VI, No. 2: 18–31.

Farlow, J. O., C. V. Thompson, and D. E. Rosner (1976). Plates of the dinosaur Stegosaurus: forced convection heat loss fins? *Science* **192**: 1123–1125.

Festinger, L., H. W. Riecken, and S. Schachter (1956). *When Prophecy Fails*. Minneapolis: University of Minnesota Press.

Feyerabend, P. K. (1975). *Against Method*. Repr., London: Verso, 1978.

Fleck, L. (1935). *Genesis and development of a Scientific Fact*. Transl. F. Bradley and T. J. Trenn. Chicago: University of Chicago Press, 1979.

Franke, R. H., and J. D. Kaul (1978). The Hawthorn experiment: first statistical interpretation. *American Sociological Review* **43**: 623–642.

Frankfurt, H., H. A. Frankfurt, J. A. Wilson, and T. Jacobsen (1946). *Before Philosophy: The Intellectual Adventure of Ancient Man*. London: Penguin, 1949.

Galbraith, J. K. 1967). *The New Industrial State*. Boston: Houghton Mifflin.

Galbraith, J. K. (1981). *A Life on our Times*. Boston: Houghton Mifflin.

Goddard, G. V. (1980). Component properties of the memory machine: Hebb revisited. In Jusczyk and Klein, Eds., pp. 231–247.

Goodman, N. (1978). *Ways of Worldmaking*. Indianapolis: Hackett.

Gould, S. J. (1981). *The Mismeasure of Man*. New York: W. W. Norton and Co.

Grey Walter, W., R. Cooper, V. L. Adridge, W. C. McCallum, and A. L. Winter (1964). Contingent negative variation: an electric sign of sensori-motor association and expectancy in the human brain. *Nature* **264**: 380–384.

Hansel, C. E. M. (1980). *ESP and Parapsychology*. Buffalo, New York: Prometheus Books.

Hayek, F. A. (1955). *The Counter-Revolution of Science*. Glencoe, Ill.: Free Press.

Hebb, D. O. (1949). *The Organization of Behavior*. New York: Wiley.

Hebb, D. O. (1951). The role of neurological ideas in psychology. *Journal of Personality* **20**: 39–55.

Hebb, D. O. (1966). *A Textbook of Psychology*. Philadelphia: W. B. Saunders.

Hebb, D. O. (1980). *Essay on Mind*. New York: L. Erlbaum Assoc.

Helmer, O. (1966). *Social Technology*. New York: Basic Books.

Hempel, C. G. (1962). Deductive-nomological *vs.* statistical explanation. In H. Feigl and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* III, pp. 98–169.

Hempel, C. G. (1965). *Aspects of Scientific Explanation*. New York: Free Press.

Hempel, C. G., and P. Oppenheim (1948). Studies in the logic of explanation. *Phil. Sci.* **15**. 135–175.

Herron, J., D. Galin, J. Johnstone, and R. E. Ornstein (1979). Cerebral specialization, writing posture, and motor control of writing in left-handers. *Science* **205**: 1285–1289.

Hesse, M. (1966). *Models and Analogies in Science.* Notre Dame, Ind.: University of Notre Dame Press.

Hessen, B. (1931). The social and economic roots of Newton's "Principia". In *Science at the Cross Roads* pp. 149–212. London: Frank Cass, 1971.

Hill, T. H. (1961). *Contemporary Theories of Knowledge.* New York: Ronald Press.

Holton, G. (1973). *Thematic Origins of Scientific Thought.* Cambridge, Mass.: Harvard University Press.

Holton, G. (1978). *The Scientific Imagination: Case Studies.* Cambridge, Mass.: Harvard University Press.

Hume, D. (1739). *A Treatise of Human Nature.* L. A. Selbidge, Ed. Oxford: Clarendon Press.

Hyman, R. (1982). Review of Alcock's book (1981). *Parapsychology Review*, March–April.

Jaki, S. (1974). *Science and Creation.* New York: Science History Publications.

James, W. (1890). *Principles of Psychology*, 2 volumes. Repr. New York: Dover, 1950.

James, W. (1907). *Pragmatism.* New York: Meridian, 1955.

Jastrow, R. (1978). *God and the Astronomers.* New York: W. W. Norton and Co.

Jevons, W. S. (1877). *The Principles of Science*, 2nd ed. New York: Dover, 1958.

Johnson, H. A. (1970). Information theory in biology after 18 years. *Science* **168**: 1545–1550.

Johnson-Laird, P. N., and P. C. Wason, Eds. (1977). *Thinking: Readings in Cognitive Science.* Cambridge: Cambridge University Press.

Kahneman, D., and A. Tversky (1973). On the psychology of prediction. *Psychological Review* **80**: 237–251.

Kant, I. (1787). *Kritik der reinen Vernunft*, 2nd ed. R. Schmidt, Ed., Hamburg: Meiner, 1930.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money.* Collected Writings Vol. VII. London: Macmillan and Cambridge University Press 1973.

Knorr, K. D. (1982). *The Manufacture of Knowledge.* Oxford: Pergamon Press.

Kripke, S. (1971). Identity and necessity. In M. K. Munitz, Ed., *Identity and Individuation*, pp. 135–164. New York: New York University Press.

Krohn, K., R. Langer and J. Rhodes (1967). Algebraic principles for the analysis of a biochemical system. *J. Computer and System Sciences* **1**: 119–136.

Kuhn, T. S. (1962). *The Structure of Scientific Revolutions.* Chicago: University of Chicago Press.

Kuhn, T. S. (1963). The function of dogma in scientific research. In A. Crombie, Ed., *Scientific Change*, Ch. 11. New York: Heinemann.

Kuhn, T. S. (1970). Logic of discovery or psychology of research? In I. Lakatos, and A. Musgrave, Eds., *Criticism and the Growth oj Knowledge.* Cambridge: Cambridge University Press.

Kuhn, T. S. (1974). Second thoughts on paradigms. In F. Suppe, Ed., *The Structure oj Scientific Theories*, pp. 459–482. Urbana, Ill.: University of Illinois Press.

Kuhn, T. S. (1977). *The Essential Tension.* Selected Studies in Tradition and Change. Chicago: University of Chicago Press.

Kuhn, T. (1978). *Black-Body Theory and the Quantum Discontinuity, 1894–1912.* New York: Oxford University Press.

Lakatos, I. (1978). *Philosophical Papers*, 2 vols. J. Worrall and G. Currie, Eds., Cambridge: University Press.

Lakatos, I., and A. Musgrave, Eds. (1970). *Criticism and the Growth of Knowledge.* Cambridge: Cambridge University Press.

Lashley, K. (1929). *Brain Mechanisms and Intelligence*. Chicago: University of Chicago Press.

Latour, B., and S. Woolgar (1979). *Laboratory Life. The Social Construction of Scientific Facts*. Beverly Hills, Cal.: Sage Publications.

Laudan, H. (1977). *Progress and its Problems*. Berkeley: University of California Press.

Lévy-Leblond, J. M. (1980). Speed(s). *Am. J. Phys.* **48**: 345–347.

Lindauer, M. (1971). *Communication Among Social Bees*. Cambridge, Mass.: Harvard University Press.

Linstone, H. A., and M. Turoff, Eds. (1975). *The Delphi Method: Techniques and Applications* Reading, Mass.: Addison-Wesley.

Lorenzen, P. (1967). Moralische Argumentationen im Grundlagenstreit der Mathematik. In *Tradition und Kritik: Festschrift für R. Zocher*, pp. 219–227. Stuttgart: F. Fromman.

Luria, A. R. (1961). *The Role of Speech in the Regulation of Normal and Abnormal Behavior*. Oxford: Pergamon Press.

Luria, A. R. (1975). Neuropsychology: its sources, principles, and prospects. In F. G. Worden, J. P. Swazey, and G. Adelman, Eds., *The Neurosciences: Paths of Discovery*. Pp. 335–361. Cambridge, Mass.: MIT Press.

Luria, A. R. (1976). *Cognitive Development. Its Cultural and Social Foundations*. Cambridge, Mass.: Harvard University Press.

Mach, E. (1905). *Erkenntnis und Irrtum*. Leipzig: Barth.

Mach, E. (1910). *Populär-wissenschaftliche Vorlesungen*, 4th ed. Leipzig: Barth.

Mair, W. G. P., E. K. Warrington, and L. Weiskrantz (1979). Memory disorder in Korsakoff's psychosis. *Brain* **102**: 749–783.

Martin, U., H. Martin, and M. Lindauer (1978). Transplantation of a time-signal in honeybees. *J. Comparative Physiology* **A124**: 193–201.

Maynard Smith, J. (1972). *On Evolution*. Edinburgh: Edinburgh University Press.

McConnell, R. A., and T. K. Clark (1980). Training, belief and mental conflict within the Parapsychological Association, *Journal of Parapsychology* **44**: 245–268.

McMullin, E. (1968). Science and the Catholic tradition. In Barbour, Ed. pp. 30–42.

McMullin, E. (1976). The fertility of theory and the unit for appraisal in Science. In R. S. Cohen *et al.*, Eds., *Essays in Memory of Imre Lakatos*, pp. 395–432. Dordrecht-Boston: Reidel.

Medawar, P. B. (1974). A geometric model of reduction and emergence. In F. J. Ayala and T. Dobzhanzky, Eds., *Studies in the Philosophy of Biology*. Derkeley and Los Angeles: University of California Press.

Merton, K. (1973). *The Sociology of Knowledge*. Chicago: University of Chicago Press.

Meyerson, E. (1921). *De l'explication dans les sciences*, 2 volumes. Paris: Payot.

Miller, G. A. (1967). *The Psychology of Communication*. New York: Basic Books.

Myrdal, G. (1969). *Objectivity in Social Research*. New York: Pantheon Books.

Myrdal, G. (1973). *Against the Stream*. New York. Pantheon Books.

Nagel, E. (1961). *The Structure of Science*. New York: Harcourt, Brace and World.

Neurath, O. (1938). Unified science as encyclopedic integration. *International Encyclopedia of Unified Science*, Vol. 1, No. 1. Chicago: University of Chicago Press.

Niosi, J. (1981). *Canadian Capitalism*. Toronto: J. Lorimer and Co.

Öpik, E. J. (1977). About dogma in science. *Annual Review of Astronomy and Astrophysics* **15**: 1–17.

Oppenheim, P., and H. Putnam (1958). Unity of science as a working hypothesis. In H. Feigl, M. Scriven, and G. Maxwell, Eds., *Minnesota Studies in the Philosophy of Science* II, pp. 3–36. Minneapolis: University of Minnesota Press.

Parkinson, C. N. (1965). *The Law and the Profits.* Hammondsworth: Penguin.

Peacocke, A. R. (1971). *Science and the Christian Experiment.* Oxford: Oxford University Press.

Peierls, R. (1979). *Surprises in Theoretical Physics.* Princeton, N. J.: Princeton University Press.

Peters, D. P., and S. J. Ceci (1980). A manuscript masquerade. *The Science* **20**, No. 7: 16–19, 35.

Pfanzagl, J. (1959). *Die axiomatischen Grundlagen einer allgemeinen Theorie des Messens.* Würzburg: Physica Verlag.

Planck, M. (1910). Zur Machschen Theorie der Physikalischen Erkenntnis: eine Erwiderung. *Physikalische Zeitschrift* **11**: 1186–1190.

Polanyi, M. (1958). *Personal Knowledge.* Chicago, University of Chicago Press.

Popper, K. R. (1959). *The Logic of Scientific Discovery.* London: Hutchinson.

Popper, K. R. (1963). *Conjectures and Refutations.* New York: Basic Books.

Popper, K. R. (1970). Normal science and its dangers. In Lakatos and Musgrave, Eds., pp. 51–58.

Popper, K. R. (1972). *Objective Knowledge.* Oxford: Clarendon Press.

Popper, K. R. (1974). Intellectual Autobiography. In P. Schilpp, Ed., *The Philosophy of Karl R. Popper.*

Popper, K. R., and J. C. Eccles (1977). *The Self and its Brain.* New York: Springer-Verlag.

Pylyshyn, Z. W. (1980). Computation and cognition: issues in the foundations of cognitive science, *The Behavioral and Brain Sciences* **1**: 111–132.

Ramage, C. S. (1976). Prognosis for weather forecasting. *Bull. Am. Meteorol. Soc.* **57**: 4–10.

Rapp, F. (1980). Observational data and scientific progress. *Studies in History and Philosophy of Science* **11**: 153–162.

Reichenbach, H. (1938). *Experience and Prediction.* Chicago: University of Chicago Press.

Rescher, N. (1973). *The Coherence Theory of Truth.* Oxford: Clarendon Press.

Rescher, N. (1979). *Cognitive Systematization.* Oxford: Basil Blackwell.

Rescorla, R. A., and A. R. Wagner (1972). A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In A. H. Black and W. F. Prokasy, Eds., *Classical Conditioning II: Current Research and Theory*, pp. 64–99. New York: Appleton-Century-Crofts.

Robinson, J., and J. Eatwell (1974). *An Introduction to Modern Economics*, rev. ed. New York: McGraw-Hill.

Rokeach, M. (1960). *The Open and Closed Mind.* New York: Basic Books.

Rorty, R. (1979). *Philosophy and the Mirror of Nature.* Princeton: Princeton University Press.

Rutherford, E. (1904). *Radio-Activity.* Cambridge: Cambridge University Press.

Salmon, W. C. (1977). A third dogma of empiricism. In R. F. Butts and J. Hintikka, Eds., *Basic Problems in Methodology and Linguistics.* pp. 149–166. Dordrecht-Boston: Reidel.

Sanders, F. (1979). Trends in skill of daily forecasts of temperature and precipitation, 1968–78. *Bull. Am. Meteorol. Soc.* **60**: 763–769.

Sawyer, J. (1966). Measurement and prediction, clinical and statistical. *Psychol. Bulletin* **66**: 178–200.

Schiff, M., M. Duyme, A. Dumaret, J. Stewart, S. Tomkiewicz, and J. Feingold (1978). Intellectual status of working-class children adopted early into upper-middle-class families. *Science* **200**: 1503–1504.

Schilling, H. K. (1973). *The New Consciousness in Science and Religion.* Philadelphia: United Church Press.

Schilpp, P. A., Ed. (1974). *The Philosophy of Karl R. Popper*, 2 vols. La Salle, Ill.: Open Court.

Schlegel, R. (1967). *Completeness in Science*. New York: Appleton-Century-Crofts.

Schlesinger, G. (1977). *Religion and Scientific Method*. Boston: Reidel.

Schlick, M. (1925). *General Theory of Knowledge*. Transl. A. E. Blumberg. New York–Wien: Springer, Verlag, 1974.

Shweder, R. A. (1977). Likeness and likelihood in everyday thought: magical thinking and everyday judgments about personality. In Johnson-Laird and Wason, Eds., 1977, pp. 446–467.

Shweder, R. A., Ed. (1980). *Fallible Judgment in Behavioral Research. New Directions for Methodology of Social and Behavioral Science*, No. 4. San Francisco: Jossey-Bass Inc.

Siberston, A. (1970). Surveys of applied economics: price behavior of firms. *Econ. J.* **80**: 511–582.

Sneed, J. D. (1979). *The Logical Structure of Mathematical Physics*, 2nd ed. Dordrecht: Reidel.

Stegmüller, W. (1976). *The Structure and Dynamics of Theories*. New York: Springer-Verlag.

Stein, P. K. (1965). *Measurement Engineering*, Vol. 1, 3rd ed. Phoenix, Arizona: Stein Engin. Serv.

Stent, G. (1978). *Paradoxes of Progress*. San Francisco: W. H. Freeman.

Suppes, P., and J. L. Zinnes (1963). Basic measurement theory. In R. D. Luce, R. R. Bush, and E. Galanter, Eds., *Handbook of Mathematical Psychology*, Vol. I, pp. 1–76. New York: John Wiley and Sons.

Taton, R. (1955). *Causalités et accidents de la découverte scientifique*. Paris: Masson.

Thompson, R. F. (1975). *Introduction to Physiological Psychology*. New York: Harper and Row.

Trevor-Roper, H. (1969). *European Witch Craze in the Sixteenth and Seventeenth Centuries*. New York: Harper and Row.

Truesdell, C., and R. Toupin (1960). The classical field theories. In S. Flügge, Ed. *Handbuch der Physik*, Vol. III/1. Berlin: Springer-Verlag.

Tuomela, R. (1980). Explaining explaining. *Erkenntnis* **15**: 211–243.

Tversky, A., and D. Kahneman (1971). Belief in the law of small numbers. *Psychological Bulletin* **76**: 105–110.

Tversky, A., and D. Kahneman (1974). Judgment under uncertainty: heuristics and biases. *Science* **185**: 1124–1131.

Tversky, A. and D. Kahneman (1977). Causal schemata in judgments under uncertainty. In M. Fishbein, Ed., *Progress in Social Psychology*. Hillsdale, N.J.: Erlbaum.

Uexküll, J. von (1928). *Theoretische Biologie*, 2nd ed. Berlin, Springer-Verlag.

Walsh, J., and D. Allen (1981). Testing the Farmer's Almanac. *Weatherwise* **34**: 212–215.

Wang, H. (1966). Process and existence in mathematics. In Y. Bar-Hillel *et al.*, Eds. *Essays in the Foundations of Mathematics*, pp. 328–351. Jerusalem: Magnus Press.

Wappes, L. (1926). Grundlegung Gliederung und Methode der Forstwissenschaft. In H. Weber, Ed., *Handbuch der Forstwissenschaft*, Vol. 1, pp. 1–42. Tübingen: H. Laupp'sche Buchhandlung.

Ward, W. C., and H. M. Jenkins (1965). The display of information and the judgment of contingency. *Can. J. Psych.* **19**: 231–241.

Wason, P. C., and P. N. Johnson-Laird (1972). *Psychology of Reasoning*. London: Batsford.

Watson, J. B. (1925). *Behaviorism*. New York: The People's Institute.

Watson, J. D. (1976). *Molecular Biology of the Gene*, 3rd ed. Menlo Park, Ca.: W. A. Benjamin, Inc.

Wheeler, J. A. (1974). The universe as home for man. *American Scientist* **62**: 683–691.

Whewell, W. (1847). *The Philosophy of the Inductive Sciences*, 2 vols. London: Frank Cass, 1967.

Wiener, N. (1948). *Cybernetics*. New York: John Wiley; Paris: Hermann.

Wiener, N. (1956). *I Am a Mathematician*. Cambridge, Mass.: MIT Press.

Williamson, P. G. (1981). Palaeontological documentation of speciation in Cenozoic molluscs from Turkana Basin. *Nature* **293**: 437–443.

Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge and Kegan Paul.

Wittgenstein, L. (1953). *Philosophical Investigations*. New York: Macmillan.

Wundt, W. (1879). Der Spiritismus. In *Essays*, pp. 342–366. Leipzig: Engelmann 1879.

Zahler, R. S., and H. J. Sussmann (1977). Claims and accomplishments of applied catastrophe theory. *Nature* **269**: 759–763.

Ziman, J. (1968). *Public Knowledge*. Cambridge: Cambridge University Press.

Ziman, J. (1979). *Reliable Knowledge*. New York: Cambridge University Press.

INDEX OF NAMES

Abraham, R. 105
Ackoff, R. 51, 231
Agassi, J. 246
Alcock, J. 108, 225
Allen, D. 48
Andreski, S. 238
Aristotle 14, 256
Asimov, I. 228
Augustine, St. 121, 236

Bachelard, G. 177
Bachrach, A. J. 110
Bacon, F. 255
Barber, B. 204
Barber, T. X. 74
Barbour, I. G. 237
Barnes, B. 261
Bechtereva, N. P. 123
Becquerel, A. 211
Békesy, G. von 174
Bell, A. G. 212
Benado, M. 89
Berk, R. A. 62
Bernard, C. 6
Bernard of Chartres 166
Bindra, D. 46–47, 187
Bitterman, M. E. 110
Blanshard, B. 150, 253
Bohr, N. 182
Bolzano, B. 116
Boole, G. 177
Boscovich, R. G. 245
Boudon, R. 72
Bourke, J. 169
Boyle, R. 233
Bridgman, P. W. 86, 95, 183, 252
Broad, C. D. 225
Brown, H. I. 261
Bube, R. H. 234
Bunge, C. 139
Bunge, M. 11, 24, 40, 75, 83, 93, 101, 130, 148, 151, 204, 215, 258, 272

Cantor, G. 177, 178
Carnap, R. 237
Causey, R. L. 32, 40, 172
Ceci, S. J. 153
Chomsky, N. 165
Claparède, E. 74
Clark, T. R. 225
Cole, S. 153
Coles, J. 111
Comte, A. 56, 252
Condorcet, Marquis de 184
Coren, S. 11
Crick, F. 7, 57

D'Abro, A. 171
Darwin, C. 44, 55, 164, 178, 182
Dawes, M. 51
de Broglie, L. 245
de Morgan, A. 177
Descartes, R. 14, 15, 45, 121, 256
Diaconis, P. 225
Durkheim, E. 62

Eberle, R. 23
Eccles, J. C. 15, 18, 74
Edison, T. A. 212
Einstein, A. 15, 139, 170, 181, 245
Eysenck, H. J. 104

Falk, R. 112
Faraday, M. 13, 164
Farlow, J. O. 111
Eatwell, J. 231
Fermi, E. 211
Festinger, L. 62

Feuerbach, L. 166
Feyerabend, P. K. 146, 165, 178, 237, 246, 252, 258
Fleck, L. 129, 175, 261
Frank, P. 150
Franke, R. H. 108
Frankfurt, H. 234
Frazer, J. 237
Frege, G. 177
Friedman, M. 231, 253

Galbraith, J. K. 62, 80
Galen 168
Galilei, G. 182, 233, 234, 255
Gall, J. 228
Gentile, G. 252
Girgus, J. S. 11
Goddard, G. V. 25
Goodman, N. 128, 150
Gould, S. J. 93
Grey Walter, W. 46
Hansel, C. E. M. 108, 225
Harvey, W. 168
Hayek, F. 14, 259, 263
Hebb, D. O. 145, 164
Hegel, G. W. F. 166
Heidegger, M. 239
Heisenberg, W. 183
Helmholtz, H. von 174
Hempel, C. G. 13, 29
Henry, J. 376
Herrenstein, R. J. 104
Herron, J. 95
Hertz, H. 77, 211
Hesse, M. 245
Hessen, B. 243
Hilbert, D. 257
Hill, T. H. 253
Holton, G. 139
Hume, D. 255
Husserl, E. 256
Hyman, R. 207

Infeld, L. 170

Jaki, S. 235
James, W. 69, 201
Jastrow, R. 15, 235
Jensen, A. R. 104
Jevons, W. S. 62
John Paul II 234
Johnson, H. A. 20
Johnson-Laird, P. N. 62

Kahn, H. 48
Kahneman, D. 47, 72, 105
Kant, I. 194, 248, 255
Kaplan, D. 26
Kaufman, W. 139
Kaul, J. D. 108
Kelvin, Lord (William Thomson) 185
Kepler, J. 233
Keynes, J. M. 152
Krohn, K. 284
Kripke, S. 32
Kuhn, T. S. 129, 152, 163, 165, 175–178, 199, 246, 261

Lakatos, I. 56, 77, 163, 176, 257
Lalande, A. 260
Land, E. 212
Laplace, P. S. 177
Lashley, K. 164
Latour, B. 242
Laudan, L. 60
Leibniz, G. W. von 15, 69, 194
Lévy-Leblond, J.-M. 95
Lindauer, M. 104
Lorentz, H. A. 170
Lorenzen, P. 252
Luria, A. R. 18, 62

Mach, E. 150, 174
MacKay, D. M. 234
Mair, W. G. P. 222
Marconi, G. 211
Marsden, J. E. 105
Marshall, A. 231
Martin, U. 104
Marx, K. 166, 178, 231
Maynard Smith, J. 177
Maxwell, J. C. 55, 164, 177, 211
McConnell, R. A. 255
McMullin, E. 163, 235
Medawar, P. 172
Mendeleev, D. 55

Merton, R. K. 160, 169
Meyerson, E. 23
Miller, G. A. 62
Montague, R. 23
Musgrave, A. 175
Myrdal, G. 161, 199

Nagel, E. 32
Neurath, O. 40
Newton, I. 25, 44, 420, 178, 233
Niosi, J. 62

Öpik, E. J. 238
Oppenheim, P. 22, 40
Oppenheimer, J. R. 55

Parkinson, C. N. 249
Paul St. 228, 236
Peacocke, A. R. 235
Peierls, R. 177
Peirce, C. S. 151, 257
Peters, D. P. 153
Pfanzagl, J. 93
Piaget, J. 5
Planck, M. 129
Plato 116, 165
Poincaré, H. 170
Polanyi, M. 129, 261
Popper, K. R. 13, 18, 70, 74, 75, 77, 79, 116, 137–138, 178, 182, 201, 245, 253
Ptolemy 55, 56
Putnam, H. 40
Pylyshyn, Z. 18

Ramage, C. S. 53
Rapp, F. 110
Reichenbach, H. 64, 70
Rescher, N. 112, 150, 151
Rescorla, R. A. 46
Ricardo, D. 152, 231
Robinson, J. 231
Rokeach, M. 63
Rorty, R. 129, 253
Russell, B. 165
Rutherford, E. 92, 99, 211

Salmon, W. C. 29
Sanders, F. 53
Sawyer, J. 51
Schiff, M. 104
Schlegel, R. 185
Schlesinger, G. 234
Schlick, M. 171
Shweder, R. A. 62
Silberston, A. 62
Smith, A. 231
Sneed, J. 163, 177
Spengler, O. 185
Stegmüller, W. 177
Stein, P. K. 93, 96
Stent, G. 185
Stevens, S. S. 92, 101
Suppes, P. 101
Sussmann, H. J. 238

Taton, R. 170
Thompson, R. F. 108
Toupin, R. 164
Trevor-Roper, H. 139
Truesdell, C. 164
Tversky, A. 47, 72, 105

Uexküll, J. von 260

Vesalius 168

Wagner, A. R. 46
Wallace, A. R. 164
Walsh, J. 48
Wang, H. 257
Wappes, L. 221
Wason, P. C. 62
Watson, J. D. 57, 174
Wheeler, J. A. 260
Whyte, L. L. 245
Wiener, N. 123
Williamson, P. G. 89
Wittgenstein, L. xvii, 127, 239, 252, 262
Wolff, C. 255
Woolgar, S. 242
Wundt, W. 145

Yerkes, R. M. 101

Zahler, R. S. 238
Ziman, J. 70, 129, 151, 201
Zinnes, J. L. 101

INDEX OF SUBJECTS

Acceptance of propositions 127
Accommodation of negative evidence 179
Account of facts 16–24
Action 285
Aims of research 200, 206–207
Anthropology of knowledge 241–242
Antireductionism 39
Automata theory 25

Background 198–200, 204–205
Basis, empirical 110
Belief 71–72
Bell's inequalities 260
Breakthrough 179
Bridge formula 32–33

Causality 25–26
Catastrophism, epistemological 179–182
Category 286
Change, epistemic 157–193
Checking 68
Chemistry 6–7, 30
Coherence theory of truth 117–118
Competition 30
Computer 187
Confirmation 132–140, 287–288
Consensus 129, 151, 200–202
Construct 116. See also Concept, Proposition, and Theory
Consistency 143–148
Context 201
Conservatism, epistemic 63
Control, experimental 285
Conventionalism 128
Cooperation 30
Correspondence principle 145
Correspondence theory of truth 117–119
Counter-revolution, epistemic 178

Data 65–68, 91–113
Deductivism. See Refutationism 95
Definition 95
  operational 86. See also Indicator
Delphi method 49
Description 5–8
Development of a field of inquiry 160–161
Dialectics 124
Differential equation 75
Discovery 64, 169–170
D-N "model" of explanation 22–24
Dogma 164, 178, 238
Domain of facts 199

Economy 211–213
Efficiency 80, 140–143
Emergence 35–37
Emergentism 40
Empiricism 70, 166–167, 253–255
Endoheresy 228
Justificationism 253
Equivalence 132
Error 67, 121–126,
Evaluation 114–154
Evidence 59–113
Evolution, biological 273–275
Evolutionism, epistemological 179–180
Exemplar 175, 176
Exoheresy 228
Expectancy 46–47
Experiment 102–111, 284–286
Experimental device 103
Explanandum 14
Explanation 5–16, 21–30
Externalism 246–247. See also Sociologism

Faith 61–62
Fallibilism 64–65, 256–257

Forecasting 45–58
Forestry 220–221
Formalism 201
Foundations research 146, 257. See also Axiomatics
Framework 176, 237–238
Freedom, cultural 251–252
Futurology 49

Generalism 174
Gestalt switch 165

Hawthorn experiment 108
History of ideas 244–247
Hypothesis
  existential 76–78
  continuity 78–79
  possibility 79

Ideology 18–19, 228–237
Incommensurability of theories 165, 181
Indicator 85–91, 143–154
Individualism, sociological 247
Innatism 165–166, 541–542
Innovation 158–165
Instrumentation 99
Integration, conceptual 41–45. See also Merger
Intuition 61
Intuitionism 61
Invention 64, 169–170

Justificationism 253, 257

Knowledge
  bogus 194
  growth of 160–161
  industry xv, 623–624, 247–248
  kinds of 194–239
  system of 215, 219

Law 206–207, 226
Level 215–216
Limits to knowledge 184–189

Magic 232
Marxism 61, 152, 231, 243
Mathematics, role of in theory construction 273–275
Meaning 164
Methodism 201–201. See also Methodolatry in Vol. 5.
Measurability 92–93
Measurement 91–102
Mechanism 171
Meliorism 65, 256–257
Merger, epistemic 41–45, 219
Message 28
Metaphor in science 4–5
Method 124–125
Methodics 200, 207
Microreduction 36–37
Mixture of fields of knowledge 219–222
Monetarism 152
Morality of science 204–205
Multidiscipline 219–222
Mystery 15

Negative evidence, disregard for 62–63
Neuropsychology 240

Objectivity 259–260
Observation 91, 284–286
Ontology 11, 85
Operationism 95
Operationalization of a theory 86–87
Originality 159–161
  degree of 282
Outlook, general 199

Paradigm 175–179
Paradogma 164
Parapsychology 225–227. See also ESP
Peer review 152
Pexgo 28. See also Psychon
Phase rule 54
Philosophy, tests of 154
Philosophy of science 147
Physicalism 39, 171
Planning 248–251
Policy of inquiry 247–252
Politology of knowledge 243–244
Postdiction 47–48
Practice 141
Pragmatism 69, 255
Prediction 47–48, 52–58
Predictor 49–50

Predictive power 282–283
Presupposition 216
Probability 29, 129, 275
Problematics 200
Prophecy 48–49
Proposal 140
Prospects of inquiry 189–191
Protoscience 194
Pseudoscience 203 223–228
Pseudotechnology 223–228.
Psychologism 167

Quantum mechanics 29, 84, 97-98, 101–102, 183

Randomness 27–28, 80
Ratioempiricism 254–255
Rationalism 69–70, 167, 253–255
Realism 70–71, 255–264
  scientific 255–258, 264–271
Reduction 31–41
Reductionism 39–41, 171–172, 217
Refutability 72, 77
Refutationism 137–140, 201
Reliability index 126
Religion 231–237
Representation 287
Research 175–179, 198–207, 221–222
Revolution, epistemic 147, 177–184
Rule 142–143

S&T system 216–219
Scenario 50–51
Science
  applied 207–209
  basic 200–207
  emergent 203
  origin of 244
  unity of 172–175
Scientism 255, 258–264
Scientometrics 161
Self-evidence 61
Semantic assumption 276–277
Semiscience 203
Simplicism 150–151
Skepticism 65. See also Fallibilism
Sociologism 167–168, 202, 242–243, 246–247
Sociology of knowledge 242–243
Solid state theory 36
Standard of measurement 99
Subject-object relation 257
Subjectivism 259–261
Subsumption 8-9, 20–21
Successive aproximations 124–125
Systematization: *See* Classification and Theory

Taxonomy. See Classification
Technology 189, 210–215, 220, 232
Teleology 30–31
Testability 63–64, 72–85
Theory 25, 262,
  choice 136–137
  merger 43–45
  reduction 33–41
  restriction 33–34
  power 288
  sequence 163
  test 82–84
Time series 57
Translation of data into theoretical language 89–90
Truth 115–129, 278-281, 287

Unanimism 69
Understanding 3-58
Unification, theoretical 31–45
Unit 98
Usefulness 129–132

Value 130–131, 148–153
Value judgment 196

W's of science and technology 58
Witchcraft 138

Mario Bunge

Treatise on Basic Philosophy

Volume 7

(Part I)

BD
161
.B86
1983
no.3
pt.1

D. REIDEL PUBLISHING COMPANY
DORDRECHT / BOSTON / LANCASTER

NUNC COGNOSCO EX PARTE

THOMAS J. BATA LIBRARY
TRENT UNIVERSITY

Digitized by the Internet Archive
in 2019 with funding from
Kahle/Austin Foundation

https://archive.org/details/philosophyofscie0007bung

TREATISE ON BASIC PHILOSOPHY

Volume 7

EPISTEMOLOGY AND METHODOLOGY III: PHILOSOPHY OF SCIENCE AND TECHNOLOGY

TREATISE ON BASIC PHILOSOPHY

SEMATICS I	1 *Sense and Reference*
SEMATICS II	2 *Interpretation and Truth*
ONTOLOGY I	3 *The Furniture of the World*
ONTOLOGY II	4 *A World of Systems*
EPISTEMOLOGY & METHODOLOGY I	5 *Exploring the World*
EPISTEMOLOGY & METHODOLOGY II	6 *Understanding the World*
EPISTEMOLOGY & METHODOLOGY III	7 *Philosophy of Science & Technology*
ETHICS	8 *The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 7

Epistemology & Methodology III:

PHILOSOPHY OF SCIENCE AND TECHNOLOGY

PART I

FORMAL AND PHYSICAL SCIENCES

D. REIDEL PUBLISHING COMPANY
A MEMBER OF THE KLUWER ACADEMIC PUBLISHERS GROUP
DORDRECHT / BOSTON / LANCASTER

Trent University Library
PETERBOROUGH, ONT.

Library of Congress Cataloging in Publication Data

CIP

Bunge, Mario Augusto.
Philosophy of science and technology.

(Epistemology & methodology ; 3) (Treatise on basic philosophy ; v. 7)
Bibliography: p.
Includes indexes.
Contents: pt. 1. Formal and physical sciences – pt. 2. Life science, social science, and technology.
1. Science–Philosophy. 2. Technology–Philosophy. I. Title. II. Series: Bunge, Mario Augusto. Epistemology & methodology ; 3. III. Series: Bunge, Mario Augusto. Treatise on basic philosophy ; v. 7.
BD161.B86 1983 no. 3 [Q175] 121 s [121] 85–2431
ISBN 90-277-1903-9 (pt. 1)
ISBN 90-277-1913-6 (pt. 2)
ISBN 90-277-1904-7 (pbk.: pt. 1)
ISBN 90-277-1914-4 (pbk.: pt. 2)

Published by D. Reidel Publishing Company,
P.O. Box 17, 3300 AA Dordrecht, Holland.

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers
190 Old Derby Street, Hingham, MA 02043, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, Holland.

All Rights Reserved
© 1985 by D. Reidel Publishing Company, Dordrecht, Holland
No part of the material protected by this copyright notice may be reproduced or
utilized in any form or by any means, electronic or mechanical,
including photocopying, recording or by any information storage and
retrieval system, without written permission from the copyright owner

Printed in The Netherlands

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy*. The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient – except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: “Do your own thing. Your reward will be doing it, your punishment having done it”.

CONTENTS OF EPISTEMOLOGY III

PART I

GENERAL PREFACE TO THE *TREATISE* v

PREFACE TO *PHILOSOPHY OF SCIENCE & TECHNOLOGY* ix

ACKNOWLEDGEMENTS xi

INTRODUCTION 1
- 1. The Chasm between S&T and the Humanities 1
- 2. Bridging the Chasm 3
- 3. Towards a Useful PS&T 5
- 4. Concluding Remarks 7

1. FORMAL SCIENCE: FROM LOGIC TO MATHEMATICS 9
- 1. Generalities 9
  - 1.1 Two Main Types of Research Field 9
  - 1.2 Some Peculiarities of Mathematics 16
- 2. Mathematics and Reality 26
  - 2.1 Conceptual Existence 26
  - 2.2 Mathematics and Reality 33
- 3. Logic 40
  - 3.1 Logic *Lato Sensu* 40
  - 3.2 Non-standard Logics 55
- 4. Pure and Applied Mathematics 75
  - 4.1 Applications of Mathematics 75
  - 4.2 An Example: Probability 86
- 5. Foundations and Philosophy 95
  - 5.1 Foundations of Mathematics 95
  - 5.2 Philosophies of Mathematics 107
- 6. Concluding Remarks 121

2. PHYSICAL SCIENCE: FROM PHYSICS TO EARTH SCIENCE 124
- 1. Preliminaries 124
  - 1.1 Physical Quantity, Convention, Measurement 125
  - 1.2 Theory, Metatheory, Protophysics 134

2. Two Classics 140
2.1 Classical Mechanics 140
2.2 Statistical Mechanics 148

3. Two Relativities 155
3.1 Special Relativity 155
3.2 General Relativity 161

4. Quantons 165
4.1 Classons and Quantons 165
4.2 The State Function and its Referents 169

5. Chance 178
5.1 Probability 178
5.2 Double Slit and Double Logic 187

6. Realism and Classicism 191
6.1 Measurement and Projection 191
6.2 Hidden Variables, Separability, and Realism 205

7. Chemistry 219
7.1 Philosophy and Chemistry 219
7.2 Is Chemistry Reducible to Physics? 226

8. Megaphysics 231
8.1 Earth Sciences 231
8.2 Cosmology 235

9. Concluding Remarks 241

BIBLIOGRAPHY 243

INDEX OF NAMES 255

INDEX OF SUBJECTS 260

PREFACE TO
PHILOSOPHY OF SCIENCE AND TECHNOLOGY

This is a systematic study in the philosophy of science and technology, or PS & T for short. It struggles with some of the so-called Big Questions in and about contemporary S & T, i.e. questions supposed to be general, deep, hard, and still *sub judice*. Here is a random sample of such problematics. Is verbal psychotherapy scientific? Is political economy ideologically neutral? Are computers creative? What is the ontological status of machines? Is engineering just an application of basic science? What is language? Are there laws of history? Which are the driving forces of history? Which is the most fruitful approach to the study of mind? Are genes omnipotent? Are species collections or concrete systems? Do the earth sciences have laws of their own? Is chemistry nothing but a chapter of physics? Does contemporary cosmology confirm theology? Has the quantum theory refuted scientific realism? Is there a viable philosophy of mathematics? How are we to choose among alternative logics? What is the ontological status of concepts?

These and other questions of interest to philosophy, as well as to science or technology, are tackled in this book from a viewpoint that is somewhat different from the dominant PS & T. An instant history of our discipline should help place our viewpoint. Modern PS & T began together with modern science and it was cultivated by scientists and philosophers until it became professionalized in the 1920s. At this time it took a *logical* turn: it was equated with the logical analysis and orderly reconstruction of scientific theories. Experimental and field work were deemed to be ancillary to theorizing, and technology was praised or deprecated, but hardly analyzed. Later on PS & T took a *linguistic* turn: only the languages of S & T seemed to matter. Facts, problems, theories, experiments, methods, designs and plans were overlooked. More recently, PS & T took a *historical* turn: everything was seen from a historical viewpoint. The logic, semantics, epistemology, ontology and ethics of S & T were declared subservient to its history or even irrelevant. Even more recently there have been attempts to force PS & T to take a *sociological* turn. Facts are said to be the creation of researchers, who would act only in response to social stimuli or inhibitors; there would be neither norms nor objective truth.

I believe the time has come for PS & T to take, or rather retake, a *philosophical* turn: to investigate the logical and semantical, epistemological and ontological, axiological and ethical problems raised by contemporary S & T, leaving the sociological and historical studies to social scientists. The time has also come to approach the problematics of PS & T in a *scientific* fashion, by paying close attention to current developments in S & T and checking philosophical hypotheses against the findings of S & T. At least this is the approach adopted in the present volume.

Although this book is part of an eight-volume treatise, it is self-contained: it can be read independently of the others. Moreover, each chapter can be read independently of the others. The book is addressed to philosophers, scientists, technologists, and culture watchers. It may be used as a textbook in a one year advanced course in PS & T. Each chapter may also be used in a course in the corresponding branch of PS & T.

To facilitate its use as a textbook, the present volume has been divided into two parts. Part I is devoted to the philosophy of the formal and physical sciences, whereas Part II covers the philosophy of the biological and social sciences as well as of the technologies.

ACKNOWLEDGEMENTS

I owe much to the many students who took my courses in PS & T: they asked interesting questions, shot down half-baked ideas, and provided valuable information. I am no less indebted to hundreds of specialists with whom I have had the privilege of discussing a host of problems in the course of four decades of scientific and philosophical research. These interactions have helped me identify and work out some of the methodological and philosophical problems that working scientists and technologists confront or skirt. They have also provided both stimulation and control.

I am particularly indebted to: my teacher Guido Beck (physics), Dave Bernardi (information technology), David Blitz (social work), Stephen Brush (history of science), George Bugliarello (engineering), Carlos F. Bunge (physics), Marta C. Bunge (mathematics), María E. Burgos (physics), Mike Dillinger (linguistics and psychology), Bernard Dubrovsky (physiology and psychiatry), Antonio Fernández-Rañada (physics), Emilio Flor-Pérez (geology), Máximo García-Sucre (chemistry), Enrique Gaviola (S & T policy), Jacobo M. Goldschvartz (physics), Ted Harrison (astronomy), Jacques Herman (sociology), Luis Herrera (astronomy), Andrés J. Kálnay (physics), Bernulf Kanitscheider (philosophy), Bernardo Kliksberg (management science), Hiroshi Kurosaki (philosophy), José Leite-Lópes (physics), Jean-Marc Lévy-Leblond (physics), Ralph W. Lewis (biology), Jean-Pierre Marquis (philosophy), Storrs McCall (philosophy), Mauricio Milchberg (information technology), Francisco Miró-Quesada (philosophy), Jesús Mosterín (logic), Jorge Niosi (economic sociology), Phineas Finn O'Jonceys (retrieving), José L. Pardos (international relations), Michel Paty (physics), Raúl Prebisch (economics), Miguel A. Quintanilla (philosophy), Osvaldo A. Reig (biology), A.C. Riccardi (paleontology), the late Jorge A. Sábato (S & T policy), Nicolás Sánchez-Albornoz (history), Yasuo Sasaki (Toyota Motor Co.), Daniel Seni (city planning), William R. Shea (history of science), Abner Shimony (physics), John Maynard Smith (biology), José Félix Tobar (engineering), Clifford Truesdell (applied mathematics), Raimo Tuomela (philosophy), Hao Wang (mathematics), Paul Weingartner (philosophy), and René Zayan (ethology). Had I listened to all their criticisms and suggestions, this would have been a better and thicker book.

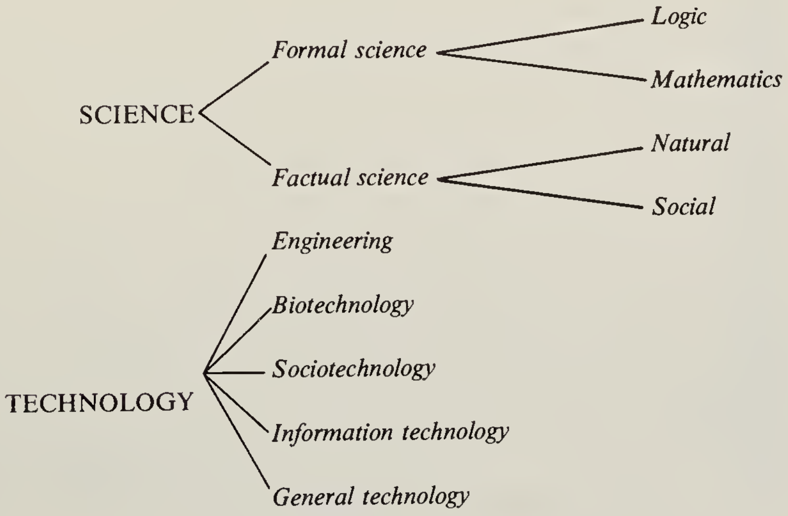
INTRODUCTION

The aims of this Introduction are to characterize the philosophy of science and technology, henceforth PS & T, to locate it on the map of learning, and to propose criteria for evaluating work in this field.

1. THE CHASM BETWEEN S & T AND THE HUMANITIES

It has become commonplace to note that contemporary culture is split into two unrelated fields: science and the rest, to deplore this split – and to do nothing about it. There is some truth in the two cultures thesis, and even greater truth in the statement that there are literally thousands of fields of knowledge, each of them cultivated by specialists who are in most cases indifferent to what happens in the other fields. But it is equally true that all fields of knowledge are united, though in some cases by weak links, forming the system of human knowledge. Because of these links, what advances, remains stagnant, or declines, is the entire system of S & T.

Throughout this book we shall distinguish the main fields of scientific and technological knowledge while at the same time noting the links that unite them. And we shall focus our attention on the following broad fields of inquiry:


SCIENCE
Formal science
Logic
Mathematics
Factual science
Natural
Social
TECHNOLOGY
Engineering
Biotechnology
Sociotechnology
Information technology
General technology

All these fields share logic and mathematics (or at least a substantial part of them) and the scientific method – which, though not used in technological design, is employed in the modeling and testing of artifacts. In addition factual science, whether natural or social, shares with technology a number of ontological, epistemological and ethical principles, such as "The world is lawful", "Hypotheses may refer to unobservables but must be checked through observation", and "Empirical information must be found out not made up".

Yet in spite of this philosophical core common to science and technology, and in spite of the fact that research in these fields continues to pose new philosophical problems, most philosophers have kept far removed from them: see Table 0.1 And only three great philosophers in the entire history

TABLE 0.1. Some influentional philosophers and their interest in S & T.

	Math.	Nat. sci.	Social sci.	Tech.
Plato	×		×	
Aristotle	×	×	×	×
Bacon				×
Descartes	×	×		
Spinoza	×	×	×	
Leibniz	×	×	×	
Locke			×	
Hobbes		×	×	
Berkeley	×	×		
Hume			×	
Kant	×	×		
Voltaire			×	
Diderot				×
Hegel			×	
Bolzano	×			
Marx	×	×	×	×
Frege	×			
Peirce	×	×		
Bergson		×		
Russell	×	×	×	×
Whitehead	×	×		
Dewey			×	
Husserl	×			
Heidegger				×
Wittgenstein	×			
Popper	×	×	×	

of mankind have paid close attention to mathematics, natural science, social science, and technology: Aristotle, Karl Marx, and Bertrand Russell. (That mathematics was not precisely a strong point of Aristotle's or of Marx's, and that Russell wrote little of lasting interest about natural science, is beside the point. We are here concerned with the interest that influential philosophers have shown in the most advanced fields of inquiry.)

Worse, the subset of influential philosophers who, in addition to having shown some interest in mathematics, science, or technology, have made any true and important statement in or about either field of inquiry, is surprisingly small. (The reader is invited to draw up his own list, and is urged not to take a nationalistic stand.) The chasm between philosophy on the one hand, and mathematics, science and technology on the other, is still huge. This makes the attempt to bridge it all the more challenging.

2. BRIDGING THE CHASM

The chasm between the humanities (in particular philosophy) and the sciences (and mathematics and technology) can be bridged, nay it is already being bridged every day in three different and mutually complementary ways. One is by rendering the humanities more scientific, as when linguistics teams up with neuroscience, psychology, sociology and history. Another is by making scientists and technologists more aware of the philosophical presuppositions and implications of their own research. A third is by engaging in the cultivation of any of the disciplines dealing with S & T.

Every science of S & T has its own concept of S & T. The psychology of S & T – a branch of cognitive psychology – treats inquiry as a mental process in an individual brain, and it may ask questions such as: What neuronal systems engage in exploratory behavior or in calculation?, or What prompted so and so to pose such and such problem? The sociology (and the economics and the politology) of S & T construes S & T as a process in a community of inquirers or, better, as a dynamic subsystem of society at large. It asks e.g. what the cultural or political stimuli and inhibitors of research are, how basic research is being applied, how best to encourage innovation, or to allocate scarce resources to individuals or teams. Finally, the historian of knowledge is supposed to treat S & T as both individual (cognitive) and social processes. He may also treat S & T as a body of knowledge, feigning that it exists separately from the brains-in-society that create, diffuse, or utilize it.

In addition to these various bridges between the humanities and S & T we have PS & T. It seems obvious that, although this discipline belongs to the humanities, it ought to make ample use of the findings of the various sciences of S & T. Unless it does, the philosopher of S & T will persist in the error of the traditional (purely internalist) philosophers and historians of knowledge, who studied knowledge as belonging to Plato's realm of ideas (or its modern equivalent, Popper's "world 3"). By proceeding in this manner the philosopher of S & T will remain estranged from the realities of S & T. He will produce descriptions of, or prescriptions for, S & T that are unrealistic and, as a consequence, will widen the chasm between philosophy and S & T instead of bridging it.

The philosopher of S & T should try and get hold of the real S & T, that which occurs in contemporary individual brains embedded in the society of today. By so doing he may do more than contribute to bridging the gap between the two cultures. He may also (*a*) enrich philosophy with a number of ideas suggested by S & T and (*b*) help S & T hone concepts and analyze propositions, spot new problems and suggest novel ways of approaching them, note links between distant fields and evaluate research projects or technological designs. By so doing he may become useful enough to S & T to earn his keep in any scientific or technological institution.

Most scientists and technologists are likely to smile, or even wince, at the idea that philosophers may tell them how to do their job. All too often they have read caricatures of scientific research or technological design. They know that real life scientists and technologists do not go about in a neat logical way but muddle through hunches and accidents, such as errors and faulty equipment. True, but most philosophers of S & T have not presumed to describe S & T but, rather, to analyze it, or reconstruct it, or discover rules for the successful conduct of inquiry. Their mistakes in these regards are no worse than some of the stories scientists occasionally tell – e.g. that the universe emerged out of nothing about 20 billion years ago, that atoms do not exist as long as they are not observed, or that the aim of science is to make careful observations.

No doubt, the good scientist (or technologist) knows how to do science (or technology) without having the logician, the methodologist, the epistemologist, or the ontologist breathe down his neck – as long as he does not come across a philosophical problem. But such an accident is bound to happen during an epistemic crisis or a revolution. And at such a juncture a philosopher, if conversant with the problem in question, may help provide analysis, criticism, or even constructive suggestion. In any case we may expect more from the cooperation of the philosopher with the scientist or the technologist than from their competition, let alone from their mutual contempt or disregard.

Philosophies of S & T are usually classed into descriptive and prescriptive. (And the former is often accused of being at best identical with history, at worst incorrect, whereas the prescriptive or normative philosophies are felt to be insolent.) However, this dichotomy is incorrect, for there are four other tasks that a philosophy of S & T may perform: analysis, criticism, rational reconstruction, and synthesis. Analysis and criticism, which are also involved in scientific or technological research, are obvious. What is not obvious is that S & T needs a dose of philosophical analysis and criticism – e.g., of their presuppositions, basic concepts or hypotheses, and procedures. As for rational reconstruction, or systematic analysis, it is the orderly reconstruction of theories and methods. Finally, philosophical synthesis deals with the general principles (ontological, epistemological, etc.) underlying the practice of S & T. In short, the various aspects of PS & T are description (of actual practice), prescription (of the best practice), analysis, criticism, rational reconstruction, and synthesis. A job for a lifetime, even if confined to a single field of inquiry.

3. TOWARDS A USEFUL PS & T

It is well known that most scientists and technologists do not listen to living philosophers. Both sides are to be blamed for this. The former for rarely reading anything outside of their own specialties, for often distrusting generalities, and for neglecting the study of the foundations of their own fields. And philosophers are to be blamed for being, as a rule, out of touch with scientific and technological advances, and for employing a jargon unintelligible to scientists or technologists.

We must try to change these attitudes that perpetuate the chasm between philosophy and S & T. The change can come only from training scientists and technologists to take a keener interest in generalities, and from reforming philosophy so that it may become attractive to researchers in other fields. In order for a philosophy of S & T to be of interest and of use to scientists and technologists, it must

(*a*) adopt the scientific approach;

(*b*) deal with real S & T, not the highly idealized and often ridiculous picture taken from popularizations;

(*c*) investigate philosophical problems that arise in the course of current scientific or technological research, or in thinking philosophically about it;

(*d*) propose precise solutions to such problems – if necessary solutions in the form of exact (mathematical) theories;

(*e*) criticize wrong approaches to research or design;

(*f*) stimulate research or design by proposing new problems, hypotheses or methods, rather than inhibit the exploration of new ideas;

(*g*) distinguish genuine from bogus S & T, deep from shallow ideas, moral from immoral policies;

(*h*) be aware of the social matrix and the history of S & T.

In short, we should aim for a PS & T that is relevant and topical, exact and systematic, critical and fertile, as well as in close touch with the other sciences of S & T.

To better indicate the kind of PS & T capable of advancing knowledge, in particular a knowledge of S & T, let us draw up a short and nearly random list of some of the problems it should investigate. This will also show that the new PS & T, far from being a chapter of epistemology, extends throughout the territory of philosophy and even somewhat beyond it.

1. *Logical problems.* What are the formal relations between two given concepts, hypotheses, or theories? What changes can we expect in a theory if the underlying logic is altered? Is it true that logic is an empirical science and may therefore be expected to change radically in the light of new scientific developments?

2. *Semantical problems.* How is factual interpretation assigned to constructs? What are the referents of the theory of evolution: species, populations, or ecosystems? How can the concept of partial truth be exactified?

3. *Epistemological problems.* What is the relation between a factual proposition and its referents? How are metastatements to be evaluated? Is the use of subjective probability ever justified in S & T?

4. *Methodological problems.* What is an indicator? What are the conceptual (in particular philosophical) presuppositions of measurement and experiment? How can the degree of confirmation of a theory be measured?

5. *Ontological problems.* How do formal objects exist? What is a law? What are artifacts?

6. *Value-theoretic problems.* What are the roles of valuation and preference in S & T? How are cognitive, practical and moral values related? Is it possible to recast decision theory in terms of objective values and probabilities?

7. *Ethical problems.* Is science morally neutral? Do technologists have moral responsibilities? Is it possible to base ethics on S & T?

8. *Esthetic problems.* Do scientists and technologists encounter any esthetic problems? Are beauty and usefulness mutually compatible? When do we say that a theory is beautiful or a proof elegant?

9. *Socio-philosophical problems.* How do ideologies affect the development of S & T? What are the philosophical underpinnings of a given S & T policy? Can philosophers do anything to change the way the public "perceives" S & T?

10. *Historico-philosophical problems.* Was it by chance that logic emerged in the first political democracy in history? Why was the study of nature reviled before the Renaissance, and exalted thereafter, in both cases *ad majorem Dei gloriam*? Why are Hume and Kant, neither of whom could read Newton's *Principia*, still widely regarded as the philosophers of the new science?

The philosopher of S & T who works on any of these or related problems, and abides by the principles laid down earlier in this Section, may do more than earn a salary and have fun: she can also be an active participant in the development of S & T rather than a distant and passive onlooker. In fact she can make contributions, to both S & T and the humanities, by:

(*a*) *digging up the philosophical presuppositions* of ideas and practices in S & T;

(*b*) *elucidating and systematizing key philosophical concepts* occurring in S & T, such as those of chance, mind, social system, meaning, truth, and explanation;

(*c*) *helping solve some scientifico-philosophical problems*, such as those of the nature of matter, life, mind, society, culture, and language;

(*d*) *proposing rational reconstructions* of research projects, scientific theories, or plans of action;

(*e*) *helping design S & T policies and plans.*

Any such work may not only advance philosophy but also S & T and even serve as a model for work in the less advanced branches of philosophy.

4. CONCLUDING REMARKS

Every expert is rightly held morally and socially responsible for the advice he offers, because his advice is part of the conceptual basis of decisions that may be beneficial or harmful to his client or to the public. The moral and social responsibility of philosophers is heaviest because their clientèle is the whole intellectual class. In fact philosophers help, for better or for worse, shape the intellectual climate of their time.

In particular philosophers of S & T, if listened to, contribute to forming attitudes – of hope or fear, love or hatred, receptivity or rejection – towards S & T. Therefore it behooves philosophers of S & T to start by becoming acquainted with the very object of their thinking, thus living up to the motto *Primum scire, deinde philosophari*. Otherwise they may promote obscurantism instead of illuminism: they may spread the idea that astronomy is no better than astrology, or evolutionary biology no truer than any creation myth; that the mind cannot be studied scientifically, or that social science is impossible; that technology is inherently evil (or good), and faith healing no worse than medicine. They may, in short, preach that "anything goes", which is indistinguishable from radical skepticism ("nothing goes").

The choice of a philosophy of S & T is a matter of great social importance in our time, when S & T lies at the very center of culture. If adopted by the people in charge of cultural policies, a wrong philosophy can maim or even kill S & T. For example, an idealist S & T policy maker would underrate laboratory and field work, as much as a positivist would overrate it. In either case the necessary balance between reason and experience would be the victim of a wrong PS & T.

Let us then try to build a PS & T both faithful to its object and capable of promoting its advancement.

CHAPTER 1

FORMAL SCIENCE FROM LOGIC TO MATHEMATICS

We start with the philosophy of logic and mathematics for several reasons. First, nowadays all of the sciences and technologies employ mathematics, which has logic built into it. Second, mathematics presupposes no knowledge of fact: it is logically, though not epistemologically, a priori. Third, for the same reason mathematics is semantically neutral: it does not describe the real world. Fourth, and consequently, mathematics is methodologically a priori: it involves no empirical procedures. In fact, the content and validity of mathematics are independent of any empirical techniques, such as finger counting and computer programming, that help carry out some calculations or construct certain proofs.

The a priori character of mathematics does not entail that it exists by itself in Plato's realm of ideas, or that mathematicians have nothing to gain from occasional contacts with science and technology. All we are stating is that, considered as a conceptual system, mathematics has no factual content and makes no essential use of empirical procedures. However, we are anticipating some results that have yet to be established. In order to obtain them it will prove convenient to begin by defining the concept of a research field. Our characterization of mathematics will result from a certain specification of this general concept.

1.1. *Two Main Types of Research Field*

Throughout this volume we shall be concerned with research fields, such as mathematics and medicine, rather than belief systems, such as religion and astrology. Since the concept of a research field will recur often in this book, we shall rehearse the definition proposed in Vol. 6, Ch. 14, Sect. 1.2:

A *family of research fields* is a set every member $\mathcal{R}$ of which is representable by a 10-tuple

$$\mathcal{R} = \langle C, S, D, G, F, B, P, K, A, M \rangle,$$

where, at any given moment,

(i) C, the *research community* of $\mathcal{R}$, is a social system composed of persons who have received a specialized training, hold strong information links among themselves, and initiate or continue a tradition of inquiry;

(ii) S is the *society* (complete with its culture, economy, and polity) that hosts C and encourages or at least tolerates the activities of the components of C;

(iii) D, the *domain or universe of discourse* of $\mathcal{R}$, is the collection of objects of study of $\mathcal{R}$;

(iv) G, the *general outlook* or *philosophical background* of $\mathcal{R}$, is composed of ontological theses (concerning the nature of the D's), epistemological principles (about the nature of inquiry into the D's), and ethical rules (about the proper behavior of the inquirers in C);

(v) F, the *formal background* of $\mathcal{R}$, is the collection of logical and mathematical theories that are or can be used by members of C, in studying the D's;

(vi) B, the *specific background* of $\mathcal{R}$, is the collection of items of knowledge obtained in other fields of inquiry and utilizable by the C's in studying D's;

(vii) P, the *problematics* of $\mathcal{R}$, is the collection of problems (actual or potential) that can be investigated by members of C;

(viii) K, the *fund of knowledge* of $\mathcal{R}$, is the collection of items of knowledge utilized by C and obtained by it at previous times;

(ix) A is the set of *aims* or *goals* of the members of C with regard to their study of D's;

(x) M, the *methodics* (usually misnamed 'methodology') of $\mathcal{R}$, is the collection of methods utilizable by members of C in their study of D's;

(xi) there is at least one other (*contiguous*) research field $\mathcal{R}'$ in the same family of fields of inquiry, such that

(*a*) *$\mathcal{R}$ and $\mathcal{R}'$* share some items in their general outlooks, formal backgrounds, specific backgrounds, funds of knowledge, aims, and methodics;

(*b*) either the domain of one of the two fields, $\mathcal{R}$ and $\mathcal{R}'$, is included in that of the other, or each member of the domain of one of the fields is a component of a system in the domain of the other;

(xii) the membership of every one of the last eight components of $\mathcal{R}$ *changes*, however slowly, *as a result of inquiry* in the same field or in related fields.

We call a research field *formal* if and only if all the members of its domain or universe of discourse D – clause (iii) in our definition – are conceptual, and *factual* if at least some of them are concrete or material. In other words, the formal/factual dichotomy is generated by the nature of the referents or objects of research fields. Thus, if a research field is about predicates in general, then it is formal, because "predicate in general" is a concept not a thing or a property of a thing. If on the other hand some of the predicates in a research field (or rather in its fund of knowledge K) are interpreted in factual terms, as is the case with "younger than", then the research field is factual even when it possesses a substantial formal background F. All of the factual sciences, pure or applied, natural or social, and all of the technologies with the exception of information technology, are factual research fields. Formal science is the only research field where all "fact" is fiction.

The radical difference between constructs and concrete things may best be seen in a few examples. *Example 1* There is no circumstance under which the identity "$2 = 1 + 1$" could fail to hold for numbers. But it may fail for things if '+' is interpreted as physical union: for example the union of two liquid droplets may result in a single droplet. *Example 2* Reference frames are representable by coordinate systems but are not identical to them. The former are things, the latter are concepts, and the former can move relative to one another, whereas the latter cannot. *Example 3* Nothingness, or the null thing, has no physical properties. However, nothing has prevented us from defining the *concept* $\square$ of null thing, which has no real counterpart, as that which adds nothing to a real thing (Vol. 3, Ch. 1, Sect. 1.1). This is not an idle fiction, for it allows us to state the general law that nothing comes out of, or ends up in, nothingness.

The formal/factual dichotomy stems from Leibniz's classical partition of the set of all propositions into *propositions de raison* and *propositions de fait* (Leibniz 1703 Bk. IV, Ch. 2, Sect. 14, and Ch. 11, Sect. 13). Such distinction is multiple: ontological, semantical, and methodological (Bunge 1984c). It is ontological because the objects satisfying a truth of reason, such as a tautology or a geometrical theorem, are constructs (e.g. propositions or conceptual triangles). On the other all the objects that satisfy a truth of fact are concrete (real, material) things. The difference is semantical because, among other reasons, an abstract formula, if satisfiable, is so in one or more models. (For example, the associative law is satisfied by the multiplication of numbers). In contrast, a factual statement, if true (to some extent), is true only *of* the only real world there is, not *in* a model (or conceptual "world").

As for the methodological differences between propositions of reason and propositions of fact, they boil down to these: (*a*) some of the former, namely the conventions (such as definitions), need not be justified except for their practical convenience, whereas all factual statements are in need of (empirical) justification: (*b*) the formal statements in need of justification are justified either by proof (deduction) or by checking satisfaction (or compliance with assumptions and definitions). And both operations are strictly conceptual. On the other hand every factual assumption calls for, at some point or other, some empirical operation – preparation, observation, measurement, or experiment – in addition to conceptual operations.

True, some mathematical propositions *seem* to be also truths of fact. For example, the theorem that the sum of the internal angles of a plane triangle equals 180° seems to hold for drawing triangles because it can be confirmed by measurements on them. This objection to the formal/factual dichotomy misses the semantical and methodological differences we have just noted. In fact, the theorem in geometry refers to a conceptual triangle, whereas the theorem in physical geometry refers to a material triangle, or rather a material thing of (*approximately*) triangular shape. And measurement *confirms*, to an excellent approximation, the theorem about physical triangles without *proving* it for conceptual triangles. No proposition is at the same time a *vérité de raison* and a *vérité de fait*. Hence both objective idealism (e.g. Platonism) and empiricism (as well as naive materialism) are wrong. More in Sect. 5.2.

We assume that pure mathematics, as classically exemplified by arithmetic, geometry, and analysis, is a formal research field, i.e. one such that all of its objects or referents are constructs, and all of its truth claims must be substantiated by purely conceptual means. This is the *conceptualist* thesis of our philosophy of mathematics, sketched in Vols. 1 and 2, and to be worked out in subsequent sections. This thesis may be traced back not just to Leibniz (and Bolzano) but also to Plato. Indeed, Plato was the first to recognize the ideal (or conceptual) nature of mathematical objects, as well as the purely conceptual character of mathematical procedures. (See Wedberg 1955.) However, we hasten to warn that our conceptualist thesis is metaphysically different from Plato's doctrine of forms or pure ideas. Far from holding that ideas exist by themselves in a realm of their own, we hold ideas to be brain processes, and constructs to be equivalence classes of such (Vol. 4, Ch. 4, Sect. 4.3., and Vol. 5, Ch. 1, Sect. 2.1.) Ours is then a *methodological* not a metaphysical dualism.

The formal/factual dichotomy, and the conceptualist thesis on the nature of mathematics, are not universally accepted by philosophers – in fact they are becoming unpopular among them. They are rejected by the radical empiricists and pragmatists, who hold that all propositions derive ultimately from experience, and all truth claims must ultimately be justified empirically (e.g. Kálmar 1967). Nor have they been admitted by such naive realists as the early Wittgenstein and the dialectical materialists, who hold that every true proposition "mirrors" or "reflects" some feature of the world. And our dichotomies have been ignored by the conventionalists, to whom all truths are stipulations or disguised definitions.

Some of the best known recent attacks on our dichotomies are those of Quine, Putnam, and Lakatos. Quine (1970) – and also Wang (1974) – rejects them in the name of the unity of "the whole interlocked scientific system". (Rejoinder: such unity is real and indispensable, but it is not harmed by the formal/factual distinction. Distinguishing does not entail detaching. The working partnership of the formal and the factual disciplines is indicated in our schema by the remark that every science has a formal background, so that it contains some mathematical propositions and methods.) Putnam (1975) claims that atomic physics has refuted ordinary logic and employs a logic of its own, where distributivity fails – which would dispose of the thesis that logic is a priori. (Rejoinder: The view that atomic physics has refuted ordinary logic derives partly from a wrong corpuscular interpretation of electron diffraction experiments: see Ch. 2, Sect. 5.2. Besides, quantum theory has ordinary mathematics built into it, and in turn the logic underlying ordinary mathematics is ordinary logic. More on this in Sect. 2.2.) Finally, Lakatos – in the wake of Gonseth, Pólya, Kálmar, and others – holds that, except for proofs, mathematical research does not differ radically from research in other fields, as it, too, draws inspiration from experience and proceeds by trial and error as well as by conjecture and counterexample. (Rejoinder: The basic methodological unity of all the sciences and, indeed, of all rational discourse and activity, is not in question and it does not prove that there are no profound semantical and methodological differences between mathematics and factual science).

We shall return to the philosophy of mathematics in several places, particularly in Sect. 5.2. Let us now see how our conceptualist thesis affects the reading of our definition of a research field. To begin with, the first two components of the ten-tuple $\mathscr{R}$ are hardly touched by our thesis on the a priori nature of mathematics, except to remind us of the ambiguity of the term. In fact 'mathematics' may be understood as denoting a *social group* (the mathematical community), the *activity* of members of such group (i.e. mathematical research), or the *product* of such activity (i.e. the changing body of mathematical knowledge). There is no harm in admitting all three acceptations of the word as long as they are not mixed up. In fact it is convenient to use them all, in order to forestall *sociologism*, which focuses on mathematics as a social group (e.g. Wilder 1981, Restivo 1983); *pragmatism*, which stresses activity at the expense of what the activity consists in and accomplishes (e.g. Wittgenstein 1978); and *idealism*, which detaches products from producers and consumers (e.g. Plato, *Republic* vi.510 C-E).

However, it may be necessary to state explicitly what the first two components of the 10-tuple $\mathscr{R}$ point to, even though it should be obvious. They suggest that mathematical research is done by members of a mathematical community hosted by a sufficiently advanced society. The community is held together by a tradition and by information flows. There is no such thing as the mathematical loner totally cut off from both the tradition and the community. And mathematical research (as distinct from the invention of a few crude and stray mathematical ideas) did not start before the dawn of civilization, and it does not prosper in contemporary backward societies. Indeed, the level of advancement of mathematics in a society is an important indicator of the level of refinement of its culture. (This is a rebuttal of Pascal's opposition between *esprit de géométrie* and *esprit de finesse.*)

The conceptualist thesis does call for a nontrivial specification of the remaining coordinates of the 10-tuple $\mathscr{R}$. To begin with, the *domain D* (or universe of discourse, or reference class) of mathematics is, according to our thesis, the collection of constructs. To be sure mathematics can be successfully applied to studying the real world. Moreover such study has often motivated the invention of mathematical ideas. Nevertheless it remains true that pure mathematics deals only with *êtres de raison*. When such constructs are employed in the study of reality, they are assigned an extra (factual) interpretation, as when a value of a distance function is interpreted as the separation between two physical objects.

Clause (iv) in our definition concerns the *general outlook* or *philosophical background* of a research field. Whatever the explicit philosophy (if any) that a mathematician may espouse, he works on a *tacit philosophy* including the following tenets. Firstly, he treats the objects of his research (functions, spaces, or what have you) as *sui generis* objects devoid of physical properties. In fact he proceeds *as if* they were self-existent in a world of their own, to the point that he prefers to say that he discovers or finds his results rather than claiming that he has constructed or invented them. Secondly, his work is theoretical: he deals with constructs, not concrete things; consequently he makes neither measurements nor experiments. Thirdly, he abides – just like the factual scientist – by a moral code that commands not to make up anything arbitrarily (except perhaps definitions), not to plagiarize, and not to keep the results of his work from others.

The *formal background F* of mathematics – clause (v) in our definition – is, trivially, mathematics itself. But of course every branch of mathematics has, at a given point in its historical development, a formal background that is only a part of the totality. For example, the propositional calculus has an empty formal background, i.e. it does not presuppose anything; on the other hand set theory presupposes the theory of deduction.

The *specific background B* of mathematics – the 6th component of $\mathscr{R}$ – is empty, i.e. mathematics need not borrow any items from other research fields. On the other hand workers in every other research field are forced to borrow bits and pieces of mathematics, which they use as tools.

The *problematics P* of mathematics – clause (vii) of our definition – is of course the collection of all possible problems, concerning constructs, that can be handled with the methods in M (the tenth coordinate of $\mathscr{R}$), and involving, at some point or other, purely conceptual (in particular theoretical) proofs or disproofs. Shorter: a mathematical problem is a problem the solution to which involves formal (i.e. nonempirical) proofs or disproofs.

The *fund of knowledge K* of mathematics – the 8th component of $\mathscr{R}$ – is the collection of results obtained previously by mathematicians and which are still valid. As is well known, the rate of decay or attrition of mathematical knowledge is the slowest. True, modern mathematicians have been able to improve on the work of Apollonius, Eudoxus, Archimedes, Euclid, and other ancient mathematicians. Still, that work stands on the whole; this cannot be said of ancient physics, with the sole exception of fragments of statics, hydrostatics, and ray optics.

The *aim A* of mathematics, or rather mathematicians – the 9th component of $\mathscr{R}$ – is of course to pose and solve mathematical problems. (The most ambitious of all such problems is the construction and inter-relation of theories.) Since mathematical problems do not concern the world – although some of them appear in the study or the control of the real world – pure mathematicians do not try to understand, let alone steer, nature or society. Only applied mathematicians, who are really factual scientists or technologists in disguise, have this other aim.

The *methodics M* of mathematics – item (x) in our ten-tuple – is composed, like that of factual science, of the general method of science plus an expanding collection of special methods or techniques. The former boils down to the sequence: *Formulate the problem-Invent hypotheses or theories-Test the solution.* The last step, testing, may consist in constructing a proof, finding a counter-example, or merely checking whether the candidate (e.g. the presumed theorem) satisfies the conditions of the problem. All such operations are exclusively conceptual, even though they may require the assistance of instruments other than the brain, such as pencils and computers. As for the special techniques, sufficice it to recall those of triangulaton and integration for the measurement of areas, and the integration of a function by expanding it into an infinite series and integrating term by term.

The *continuity* condition (xi) is satisfied both locally, i.e. by every branch of mathematics, and globally, by mathematics as a whole. In fact there are no isolated chapters of mathematics: the latter makes up a conceptual supersystem. Consequently every mathematical specialist can talk to colleagues in at least one other field of mathematics. Moreover, mathematics interacts vigorously with factual science and technology: it lends them formal tools, and receives in exchange new problems and new tools – namely calculators and computers. Mathematics is thus at the very center of the science and technology system: though conceptually self-contained, it is not self-sufficient, let alone selfish. This is why those who speak the language of mathematics can talk to specialists in all the advanced sciences and technologies. Only philosophers can compete with mathematicians in universality – particularly if they understand the universal language.

Finally, the *changeability* condition (xii) is met automatically as long as there are active research mathematicians, because the job of every investigator is to innovate in his field, not just to teach it or use it. Thus it is believed that ancient mathematics died when Hypatia of Alexandria was stoned by a christian mob. There was almost no mathematical creation in Medieval Western Europe: there was only the occasional study of a few Greek mathematical texts, usually misunderstood. Original research was not resumed vigorously until the Renaissance: for more than one millennium mathematics had shrunk and degenerated from a research field into a belief system. The same may happen again if the current suspicion of pure science prevails.

So much for our general characterization of mathematics. Let us now study in some detail some of the peculiar traits of mathematics, i.e. those that mark it off from other research fields.

1.2. *Some Peculiarities of Mathematics*

Mathematics can be studied from within or from without. The internal study of mathematics can proceed on two levels: it can bear on mathematical objects, such as sets or algebraic systems, or on propositions about such objects. In the former case one conjectures or proves mathematical statements, in the latter one deals with metastatements, i.e. statements about statements. For example, the formulas "$p \vee p \Leftrightarrow p$" and "$p \& p \Leftrightarrow p$" are mathematical (more precisely logical) propositions, whereas the statement that they are duals of each other is a metastatement, and so it belongs in metalogic. Likewise, when one tries to prove the consistency of a given abstract theory by exhibiting a model (example) of it, one engages in metamathematical research. The same holds for studies of completeness, categoricity, decidability, and the like.

We shall make only passing remarks on metamathematics, which is part of mathematics – just as a philosophical reflection upon philosophy is itself part of philosophy. Metamathematics is secure in the hands of mathematicians, who have advanced it tremendously in the course of its short history. (See e.g. Kleene 1952, Rasiowa and Sikorski 1970, Hilbert and Bernays 1968, 1970, Barwise ed. 1977.) We shall be far more interested in the philosophy of mathematics, which is one of the various external studies of mathematics – others being the psychology, sociology, and history of mathematics. Unlike metamathematics, which is hardly one century old and yet fairly advanced, the philosophy of mathematics, though at least 25 centuries old, is still underdeveloped. This is how two mathematicians interested in the subject have recently evaluated it: the philosophy of mathematics, is "dormant since about 1931" (MacLane 1981), it "has not progressed beyond 1931" (Smorynski 1983) – 1931 being the year Gödel published his incompleteness theorems.

The philosophy of mathematics is mostly a collection of interesting open problems and of ill-grounded opinions. Moreover, it is fragmentary rather than systematic. What is worse, it is quite isolated from the other branches of philosophy; in particular, none of the well known philosophies of mathematics is a component of a philosophical system consonant with contemporary mathematics and factual science. There is not even consensus on the problematics of the philosophy of mathematics. Thus a well-known reader (Hintikka 1969) sidesteps all the classical questions on the nature of mathematics, to concentrate on technical problems in the logical foundations of mathematics. In sum, the philosophy of mathematics is in poor shape. Therefore it poses one of the most interesting and pressing challenges to any philosopher who is aware that the most advanced and exact of all sciences has the most backward, woolly and dogmatic of all philosophies.

In this section we shall grapple with a few basic problems in the philosophy and psychology of mathematics, namely: What do mathematicians typically do?, How do they proceed: what kind of methods do they use?, What is the universe of discourse (reference class) of mathematics?, What is the meaning of mathematical constructs?, What is mathematical truth (in contrast to factual truth)?, What is the difference between pure and applied mathematics?, and Do mathematicians discover or invent?

What do mathematicians do? They are supposed to think hard about conceptual problems of certain kinds. More precisely, mathematicians carry out computations (e.g. integrations) or invent new algorithms (e.g. schemes of numerical computation); they hypothesize or prove theorems in existing theories; they introduce new concepts, theories, or even entire branches of mathematics; they organize (e.g. axiomatize) known bodies of mathematics; they simplify algorithms, proofs, or theories; they analyze (in particular criticize) their own work or that of others; they investigate global properties of theories (e.g. completeness); they help colleagues in science or technology solve some of the mathematical problems that occur in them – and much else. All such tasks are strictly conceptual.

The nature of such tasks determines the way to go about them. Fingers, pencils, rulers, and computers can help, but they cannot do mathematical work because they are brainless. For example, a computer cannot work in number theory because it works with numerals not numbers, figures not geometrical objects, symbols not what these stand for. (For the difference between constructs and the signs or symbols that represent them see Vol. 1, Ch. 1.)

It has become fashionable to claim that mathematics is not fundamentally different from factual science because mathematicians often proceed by trial and error and they may employ material tools such as compasses and calculators. (E.g. Lakatos 1976, Davis and Hersh 1981.) Actually all this shows is that mathematicians are not disembodied spirits but whole organisms whose brains include sensorimotor systems connected with the rest of the body. It also shows that a mathematician uses a single brain whether to prove a theorem or solve some practical problem. Like anyone else, the mathematician is bound to use analogy and induction, and to try conjectures until hitting on the correct solution. (See Pólya 1954 for a beautiful array of examples of nondeductive reasoning in mathematics.)

If we wish to find out what sets mathematics off from other research fields we must look for differences not for similarities. An extremely simple example will suffice to show where the difference lies. Suppose we wish to prove that, for any quadruple of real numbers a, b, A, and B, $aA = bB$ if, and only if, $a/b = B/A$. An ordinary mathematician solves this problem by recalling the definition of number division in terms of number multiplication, namely: "For any real numbers x, y, and z, $x/y = z =_{df} x = yz$". An empiricist (and also a vulgar materialist) philosopher might propose, instead, studying the behavior of levers of arm lengths a and b, and respective weights A and B, in equilibrium states. By measuring these quantities he would in fact find that his results satisfy the given theorem, which is structurally identical with Archimedes' law for balanced levers. But of course he has only *exemplified* the theorem instead of proving it. (And even exemplifying the theorem in a physical case will require rounding off the measurement results so as to mask the errors.) He has not proved that the theorem holds for infinitely many combinations of real numbers a, b, A, and B. (See Perdang 1983 for an optical "proof" of Brouwer's fixed point theorem.)

In short, mathematicians are human beings and, like all humans, they employ certain thought patterns common to all human beings; they are also capable of drawing inspiration from activities in other walks of life. But they train themselves to think of constructs in general terms. To a mathematician an example may suggest looking for a pattern or refuting it. An accumulation of examples (i.e. what would constitute substantial evidence in factual science) is no substitute for a proof. Moreover, just as in science, individual cases only pose the problem of finding the underlying patterns characteristic of a system of constructs. Mark the words *system* and *pattern*, because they are the clues to our next question: What is mathematics about?

Historians of mathematics have noted that, until about mid-nineteenth century, the bulk of mathematical research was concerned with individual constructs, such as particular figures, equations, functions, or algorithms. From then on, and particularly since mid-20th century, mathematics has been conceived as the study of *conceptual systems*, such as groups of transformations (or even the whole category of groups in general), families of functions (or even entire functional spaces), and topological spaces (such as metric spaces in general). (Caution: Bourbaki, Bernays and others call 'structure' what others, e.g. Hartnett 1963, call 'system'. We stick to our convention in Vols. 3 and 4 that every structure is the structure *of* some object: that it is the set of all the relations among its components – the internal structure – plus those among the latter and the environment or context of the system, which can be empty – the external structure.)

Every mathematical system ("structure") can be characterized in either of two ways: (a) as a set equipped with a structure consisting of one or more operations or functions defined on that set (e.g. Bourbaki 1970); (b) as a collection of objects together with one or more morphisms relating those objects – i.e. a category (e.g. MacLane 1971). (Actually the second concept subsumes and supersedes the first.) We distinguish two kinds of mathematical system: (*a*) *concepts*, such as that of the Euclidean plane, or that of the family of all the circles on that plane, and (*b*) *theories* (hypothetico-deductive systems) about conceptual systems in sense (*a*) – e.g. plane Euclidean geometry.

Example 1 The *system of natural numbers*, as specified in part by Peano's five axions, is $\mathcal{N} = \langle \mathbb{N}, 0, ' \rangle$, where $\mathbb{N}$ designates the set of non-negative integers, 0 a distinguished element of $\mathbb{N}$, and $'$ the successor relation. The latter generates all of the members of $\mathbb{N}$ out of 0 – e.g. $2 = 0''$. The conceptual system $\mathcal{N}$ is the subject or referent of another conceptual system, namely elementary number theory. *Example 2* A *semigroup* is a system of individuals glued together by a binary associative concatenation operation $*$: $\mathcal{S} = \langle S, * \rangle$, where S, the base set of $\mathcal{S}$, is an arbitrary (abstract) set. The entire set S may happen to be generated by concatenating a single individual a with itself a number of times, i.e. by forming $a*a$, $a*a*a$, etc. *Example 3* A *theory* is a system $\mathcal{T} = \langle P, \vdash \rangle$, where P is a set of propositions including logic, and $\vdash$ the relation of (syntactic) entailment, or logical consequence. In our systems-theoretic terminology, P is the *composition*, and $\vdash$ the *structure* of $\mathcal{T}$. A theory is *formal* if all the referents of the members of P are conceptual, and *factual* if some of them are concrete (material, real). If $\mathcal{T}$ is factual, then P contains a set of *semantic assumptions* ("correspondence rules") that specify the factual content of $\mathcal{T}$. These semantic assumptions are of the forms "*c refers to f*" or "*c represents f*", where c is a construct occurring in some P's, and f is a factual item, such as a material thing, a property (or state) of it, or an event (change) in it. There are no such assumptions in pure mathematics. On the other hand a theory in applied mathematics is characterized by a non-empty set of semantic assumptions.

Talk about semantics is essentially talk about meaning or truth. What is the meaning of mathematical constructs? In our semantics (Vols. 1 and 2) only predicates, such as "$\in$", and propositions, such as "$1 = 0$", have meaning. Hence sets are devoid of meaning even when their members are meaningful. But of course sets allow one to construct objects endowed with meaning, such as the predicate which, to each ordered pair $\langle x, y \rangle$, with $x \in \mathbb{R}$ and $y \in [-1, 1]$, assigns the proposition "$\sin x = y$". Meaning, then, can emerge by combining meaningless concepts, or it can be submerged by collecting meaningful concepts into sets. E.g., numbers are meaningless because they are neither predicates nor propositions.

But what is meaning? This question is particularly important in view of the fact that some complaints against standard mathematics boil down to the accusation that it is full of "meaningless" concepts and even propositions. (See e.g. Bishop 1967, Stolzenberg 1970.) Unfortunately those critics do not tell us what they mean by "meaning". In our semantics the meaning of a construct in a given context equals its sense together with its reference. (Vol. 2, Ch. 7.) In turn, the sense of a predicate or formula in a given context (e.g. theory) equals the set of all the predicates (or formulas respectively) that are logically related to it, i.e. that entail it or follow from it. As for reference, let us recall that we construe a predicate as a function pairing individuals, or *n*-tuples of such, to the propositions containing the predicate. Those individuals are precisely the referents of the predicate. More exactly, the reference class of a predicate is the union of all the sets occurring in its domain. (For example, the reference class of "give" is the union of donors, acceptors, and things given.) And the reference class of a proposition is the union of the reference classes of the predicates occurring in the proposition. (For example, the reference class of "All mathematicians love some theories" is the union of the class of mathematicians and that of theories.) In sum, the *meaning* of a predicate or proposition p in a context (e.g. theory) T is

$$\mathscr{M}ean_T(p) = \langle \mathscr{S}_T(p), \mathscr{R}_T(p) \rangle .$$

Note the following points. First, meaning is contextual. Thus "triangle" does not have the same sense in synthetic geometry as it does in trigonometry, which – unlike the former – is a metric geometry, even though both have the same referents. Likewise "set" has different meanings in alternative set theories – i.e. these define different notions of set. This systemic view of meaning is at variance with both the semantical individualism upheld by Frege and Brouwer (every proposition has an intrinsic meaning) and Quine's semantical holism (the meaning of a construct is determined by the totality of theories, formal or empirical, where it occurs).

Second, in keeping with the conceptualist thesis adopted in Sect. 1.1, we say that the reference class of a mathematical predicate or proposition is a set of conceptual objects, i.e. of objects belonging to some conceptual system in either of the two senses elucidated a moment ago. For example, plane trigonometry refers to all the triangles on the Euclidean plane – mathematical triangles to be sure, not physical ones formed e.g. by intersecting sticks or laser beams.

A moment ago we stated that mathematics studies conceptual systems ("structures"). However, this is only a necessary condition: philosophy and the history of ideas too study conceptual systems, such as cosmologies and mathematical theories. What makes the mathematical study of conceptual systems unique is that (*a*) it is *purely conceptual* (i.e. does not make essential use of any empirical data or procedures) and it involves, at some point or other, (*b*) *positing or conjecturing the laws* (general patterns) satisfied by the members of those conceptual systems, as well as (*c*) *proving or disproving conclusively* some such conjectures (Mathematical proofs are perfectible, but they are conclusive.) There are shorter and more beautiful definitions of mathematics, e.g. "mathematics is the *science of the infinite*" (Weyl 1949 p. 66). However, mathematics is too complex a subject to be adequately characterized by a single short phrase; for example, Weyl's definition overlooks finite mathematics.

(Rigorous proof by deduction, unlike rigorous refutation by counterexample, is distinctively mathematical. For this reason proof is often regarded as the heart of mathematics. But this is like saying that the heart of physics is hypothesis testing. Checking – proving theorems and disproving conjectures – may certainly take up most of the time of a mathematician, but it is not more important than looking for new problems, restating old problems in more suitable terms, conjecturing theorems or algorithms, building theories, or using them for solving problems. After all, most theorems must be hypothesized before they can be proved, and usually the motivation for conjecturing them is the need to solve some problem. In short, proof is certainly distinctive of mathematics, but there is more to doing mathematics than proving or disproving.)

We may then define *contemporary pure mathematics* as the investigation, by conceptual (a priori) means, of problems about conceptual systems, or members of such, with the aim of finding (inventing or discovering) the patterns satisfied by such objects – a finding justified only by rigorous proof. On the other hand we define *contemporary applied mathematics* as the investigation of problems arising in factual science, technology, or the humanities, with the help of constructs belonging to pure mathematics. (Occasionally the applied mathematician may have to supply such constructs himself, but to him they are means, not ends.)

Applied mathematics is then distinguished from pure mathematics by (*a*) the *source* of problems, which is extramathematical in the former case and internal in the latter; (*b*) the *ultimate referents*, which are real things in the case of applied mathematics, and constructs in the other case, and (*c*) the *aim*, which is to help some nonmathematical discipline in the first case, and to advance pure mathematics in the second. (More on the pure/applied distinction in Sect. 4.)

In characterizing mathematical research we have stressed that it involves proving. Now, a proof is a valid reasoning, or argument, from a set of premises to a conclusion, with the help of rules of deductive inference such as, e.g., the modus ponens and universal instantiation. However, not all proofs are alike: whereas some are rigorous, others are informal. And whereas some theorems are provable in a manageable number of steps, i.e. within a lifetime, others are not. In other words, there are *degrees of rigor*, hence of provability or theoremhood. Mathematicians reach for maximal rigor but sometimes they have to settle for less. In such cases they may find consolation in the fact, well known to historians, that rigor, far from being absolute, is time-bound. Thus, although Euclid's theorems are correct, all of his proofs of them can be improved.

Whatever the rigor of the proof of a formula, we may say that the proof *justifies* the formula. We may also say that the proof shows the formula to be *formally true*, or a *truth of reason*, to distinguish it from a truth of fact. As Grassmann (1844, Einleitung) noted long ago, whereas the factual (or "real") sciences find truth in the correspondence between thought and external objects, formal truth consists in the matching among reasoning processes: it is a thoroughly internal affair.

But if we wish to avoid confusing truths of reason with truths of fact – a confusion generating the illusion that a single theory can account for both – we may as well dispense with the notion of formal truth. (The illusion was dispelled in Vol. 2, Ch. 8, Sect. 2.) This tactics is advocated by a number of mathematicians (e.g. Bourbaki 1970) and philosophers – namely the fictionists or conventionalists. It is part and parcel of the conventionalist philosophy of mathematics, which boils down to the thesis that all mathematical formulas are conventions, so that they can be rich or poor, deep or shallow, useful or useless, elegant or clumsy, but never true in the same sense as some of the formulas or chemistry or sociology. We shall return to this subject in Sect. 5.2. Suffice it to say here that the non-truth view of mathematical truth contains a grain of truth.

The last problem in our agenda in whether mathematical objects are discovered or invented. Idealists, such as Plato, Bolzano, and Frege, hold that all ideas are already in existence somewhere, so all we can do is discover them – with hard work and luck to be sure. One good reason for holding this view is that, when constructing a mathematical object, or a theory concerning such an object, we need not posit all of its properties. In fact, when constructing a concept we need posit only its generating properties: we discover the rest as we study it. Likewise when constructing a theory we need postulate only some propositions: the rest can be deduced as we work out the theory; or, if any additional postulates are needed, we can supply them as we proceed. In short, we invent the assumptions and discover their consequences.

So, there is evidence that mathematicians do discover. Moreover we can explain how this comes about, namely with the help of our previous assumption that mathematics is about (conceptual) systems. Indeed, it is characteristic of systems, unlike mere aggregates, that their components and properties are interdependent (lawfully inter-related). Hence, knowing some of them we can infer (discover) the others. This explains the well known fact that some mathematical results are discovered independently by several investigators. But it does not explain why others are the brain children of a few exceptional individuals.

The evidence for mathematical discovery does not suffice to justify Platonic idealism, for the following reasons. First, before we can discover that A entails B, someone must have posited A, and such positing is an invention, not the discovery of something that is already "there". Second, there is no evidence for the thesis that ideas, like mountains, are self-existing. On the contrary, there is plenty of evidence from physiological psychology that ideation is a brain process. (Recall Vol. 4, Ch. 4, or Vol. 5, Ch. 1.). Third, if mathematical ideas were already "there" (where?), up for grabs, mathematical research could proceed in a social vacuum, since productivity would depend only on individual talent. But this is not so: mathematical research is always done in some community of inquirers who inherit a cultural tradition, because mathematicians need to learn from others and they need to be stimulated and controlled by others. (Recall Sect. 1.1.)

In conclusion, in mathematics there is invention as well as discovery. We invent when positing (assuming), and we discover when finding out what our posits commit us to (logically). Think of Cauchy's famous integral formula: the value of an analytic function at an arbitrary point inside a region bounded by a closed continuous curve is determined by the totality of values of the function on the boundary: See Figure 1.1. That is, once the border and the values of the function on it have been posited, there is no room left for further invention: the interior is rigidly determined by those boundary conditions. All we seem to be doing, when using Cauchy's formula to compute a value of the given function at an interior point, is to discover some eternal Platonic idea dwelling in that domain. However, such a discovery is still a creation: the newly calculated value was *nowhere* before

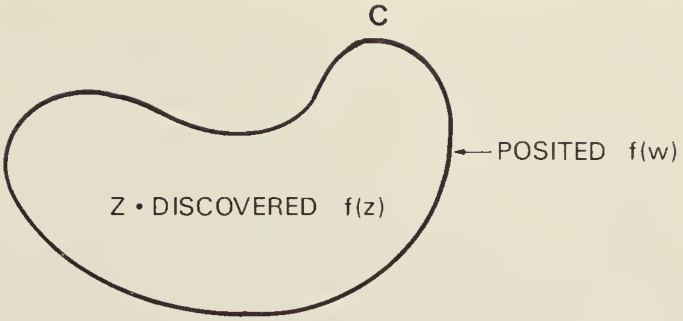


Fig. 1.1. The Cauchy integral formula. Positing all the values of an analytic function f on the curve C determines uniquely the value of f at an arbitrary interior point z:

$$f(z) = (1/2\pi i) \int_C \frac{f(w)\,\mathrm{d}w}{w - z}$$

The positing is an invention, the calculation a discovery, and both are creative acts. Moreover they are original creations, though possibly modest ones, if nobody else performed them previously.

it was computed. All there was is a network rigidly determined by the boundary conditions: these are the generating components of the network and the only ones known beforehand. The remaining components are discovered thanks to that marvelous invention, Cauchy's integral theorem.

Whether invented or discovered, an original mathematical contribution is a *creation*: it brings to light something we did not know before. (Recall Vol. 5, Ch. 6, Sect. 1.2 on the creative power of deduction.) Of course, logicians tend to think that all the infinitely many components of a theory are born the moment the postulates are laid down. But this is just a *façon de parler*: a theorem that has not been proved, let alone guessed, is nowhere: it does not exist except potentially. What happens is that, for mathematical and philosophical purposes, we must *pretend* that such objects exist (formally), i.e. that they are components of a theory known only partially to us. In other words, we must distinguish between a construct and our knowledge of it, and feign that the construct exists (formally). A few examples will show why the distinction and the pretense are necessary – *pace* mathematical intuitionism.

Example 1 A hypothetico-deductive theory is composed of infinitely many propositions: logic "generates" them all out of a finite set of premises. We can get to know only a finite subset of the totality of formulas of a theory. Yet if we know the postulates of the theory we can study certain global properties of it, such as consistency and completeness. *Example 2* A formula that holds for every member of an infinite set may be assumed or proved even though nobody is capable of thinking (hence knowing) every single instance of the formula. *Example 3* (Fréchet 1946). A rational number is defined as the ratio of two integers even if, given a particular number, we may be unable to find two integers that determine it (as a rational number). Thus we distinguish definitions from criteria, tests, and rules. Otherwise we would have to admit either that there are numbers which are neither rational nor irrational, or that there are gaps in every set of rationals.

In short, when doing mathematics (or metamathematics) we proceed *as if* we were Platonists: we feign that mathematical objects, and the formulas about them, exist on their own. This pretense entails our distinguishing mathematical statements from knowledge claims or disclaims, such as "We know how to construct, prove, or recognize x" – which is about knowing subjects and therefore belongs in epistemology or psychology. However, we are already encroaching on the territory of the following section.

2. MATHEMATICS AND REALITY

2.1. *Conceptual Existence*

Throughout this *Treatise* we have distinguished constructs, such as concepts and propositions, from factual items, such as concrete things and their states. Thus we have held that constructs have peculiar – mathematical and semantical – properties that factual items lack. (See e.g. Vol. 1, Ch. 1.) But at the same time we have contended that all constructs are created by rational animals and, more precisely, that they may be construed as equivalence classes of brain processes of a certain type. (See e.g. Vol. 5, Ch. 1.) Therefore our construct/fact dichotomy is methodological, not ontological. That is, we postulate that there is but one real world which is composed exclusively of factual (concrete, material) entities, some of which – people among them – are capable of creating constructs. (For the notion of a construct see Vol. 1, Ch. 1, Sects. 1.2 and 3.1; for mind-body monism, see Vol. 4, or Bunge 1980a, Ch. 5; for the construal of a construct as an equivalence class of brain processes, see Vol. 5, Ch. 1, or Bunge 1981a.)

Now, distinguishing constructs from factual items amounts to distinguishing *formal* (conceptual, ideal) *existence* from *factual* (concrete, material) *existence*. Thus we say that, whereas numbers exist formally, electrons exist factually. One way of elucidating this difference is by stating that, whereas numbers are members of the collection of constructs, electrons belong in the collection of material objects. An equivalent way is to introduce two different existence *predicates*: conceptual (or formal) and material (or real). These predicates, rather than the allegedly all-purpose "existential" quantifier $\exists$ of logic, should be employed when discussing ontological matters. (See Vol. 3, Ch. 3, Sect. 4.3, or Bunge 1981a.) Unless these two existence concepts are distinguished and rendered precise, expressions such as 'there are no abstract entities' (found e.g. in Field 1980) are so ambiguous that they make hardly any sense. Worse, failure to draw the distinction may lead to asserting that mathematics and science are about nonexistents (Routley 1980).

Our next task is to characterize more precisely the notions of real and formal existence. We take it that the keys to real existence are *absoluteness* and *mutability*, whereas those to formal existence are *context-dependence*, *immutability*, and *conceivability*. Concerning absoluteness: whereas real things exist absolutely, every construct exists in some context or other, e.g. by fiat or by proof in some theory. For example, the natural numbers exist (formally) in number theory but not in lattice theory. On the other hand electrons and cells, mountains and forests, brains and societies, do not exist relative to some context only to vanish in another: they exist absolutely. (Caution: although things exist absolutely, certain properties, and consequently the changes in them, are frame-dependent. See Vol. 3, Ch. 5, Sect. 2.5.)

As for mutability, the difference between the two types of object is this: whereas material things are changeable, constructs are not. Thus photons keep moving as long as they exist, whereas numbers do not budge. Therefore we attempt to establish the true equations of motion for photons, not for numbers. (The equations themselves, being constructs, do not change. But of course we do change when thinking of alternative equations.) The converses too are true: whatever is changeable exists really (materially), and whatever is immutable exists only formally (conceptually). True, the derivative of a function is sometimes described informally as "the rate at which the function is changing". But this is only a hold-over from 18th century mathematics. Functions do not change although they may be made to represent changing properties of concrete things.

(To put it in the state space language adopted in Vols. 3 and 4: whereas material objects can be in at least two different states, formal objects are always in the same state – which is tantamount to saying that there is no point in assigning them a state. Shorter: material objects can be assigned state spaces, formal objects cannot. See Bunge 1981a.)

Mathematical intuitionists agree that mathematical objects do not exist by themselves but come into existence as they are intuited or constructed. But they add that, since intuition and construction are processes, what they produce is just as temporal as a material object. (This is why they reject the concept of actual infinity as unintelligible: "we can arrive at the notion of infinity in no other way than by considering a process of generation or construction which will never be completed": Dummett 1977 p. 56.) That mathematical research is a process, is obvious but irrelevant to the question of the nature of its products. We must distinguish (though not separate) the activity from its product, e.g. research from its finding, because the former is personal whereas the latter may be universal. (Recall Vol. 5, Ch. 2 on the distinction between cognition and knowledge.) Moreover in standard ("classical") mathematics we pay no attention to the psychology of research and *feign* that all the admissible mathematical objects are ready made: we thus distinguish mathematics from the creation or learning of it. There is nothing to prevent us from inventing such a fiction precisely because mathematical objects are *entia rationis* quite unlike material objects, which are forever in a state of flux and satisfy factual (e.g. physical) laws. (More on mathematical fictionism in Sect. 5.2.)

To put it into other words: When creating or discovering mathematical objects we fashion them *as* atemporal items. In other words, mathematical propositions and, in general, truths of reason – unlike truths of fact – are tenseless: i.e. the concept of time fails to occur in them. Thus we distinguish the mathematical proposition "π is transcendental" from the factual proposition "π was shown to be transcendental one century ago". If intuitionists were consistent they would introduce the concept of time as a basic (primitive) concept of mathematics. (As a matter of fact once Brouwer did try do do just this.) But this would lead them into contradiction, for such a concept must be elucidated with the help of prior notions, such as those of set and function, not to speak of those of event and reference frame. (Recall Vol. 3, Ch. 6, Sect. 3.) By avoiding this inconsistency, intuitionists behave inconsistently.

Finally, the conceivability condition is this: for something to be a construct it is necessary and sufficient that it may be conceived by someone. (On a realistic epistemology like ours it is not necessary for a real thing to be conceivable in order to exist. The universe existed before it contained animals capable of conceiving it. On the other hand there were no mathematical objects before the emergence of animals capable of conceiving them.) In other words, we stipulate that x is a *construct* if, and only if, there exists (really) an animal capable of conceiving x as a conceptual system or a member of such. (Recall our definition of a conceptual system in Sect. 1.2.)

Our preceding convention allows not only for mathematical constructs but also for mythological ones. We are far more demanding when it comes to mathematical existence. We demand that it be defined in or by a consistent theory dealing with exact concepts. It follows that the constructs typical of fuzzy or inconsistent theories are devoid of mathematical existence: they do not differ essentially from mythological characters. (Constructivists are even more demanding: for them only constructive existence, and moreover provided it is ultimately reducible to positive integers, is legitimate. See Sect. 5.) We shall come back to this point in a short while.

In sum, whereas real existence is absolute (context-free) and is characterized by mutability, formal existence is relative (context-dependent) and is characterized by immutability and conceivability. Hence in our view Plato was right in holding that ideas are immaterial and immutable, wrong in believing that they exist really by themselves. And Hegel was wrong not only in asserting that ideas are self-existing but also in holding that they are characterized by self-activity: that they go by themselves through the three dialectical stages (thesis, antithesis, synthesis) propelled by contradiction. Only real (material) entities are changeable.

So much for formal and material existence in general. Let us now turn to particular existence statements, such as "There are infinitely many lines passing through any given point in three-dimensional space". Such existence statements are just as essential in mathematics as real existence statements are in factual science. In the latter we certify or at least conjecture explicitly the existence of entities of a certain kind before proceeding to accounting for them in any great detail. (Characteristically, ideology and pseudoscience describe objects the existence of which has not been or cannot be established.) And in ordinary mathematics one postulates or proves the existence of objects of a certain kind before proceeding to conjecture or prove the laws they satisfy. (In constructive mathematics one must start by constructing the objects concerned with the help of explicit procedures such as computations.) Indeed, the antecedent of a general mathematical law is (tacitly or explicitly) an existence statement, omitted only when taken for granted. The simplest such law is of the form: "If x exists, then: If x has property F, then x has also property G". Omission of the cautious existence clause may result in laborious manipulation of ghosts. (If existence is not presupposed or proved, some system of free logic must be assumed, or else the formal existence predicate introduced by Fourman 1977 must be used.)

Every existence statement can be construed as a statement about the membership of the object(s) concerned in some set, class, or category specified by an exact consistent theory. *Example 1* The empty set axiom, namely "There is a set without members", is replaceable with: "The category of sets contains one set without members". *Example 2* Euclid's theorem about the existence of infinitely many prime numbers can be recast thus: "The set of prime numbers is an infinite subset of the set of natural numbers". *Example 3* Any theorem asserting the existence of real functions solving differential equations of some type can be translated into the assertion that the set $\mathbb{R}^{\mathbb{R}}$, where $\mathbb{R}$ is the real number system, contains functions satisfying the given equation.

Our thesis can now be restated as follows:

DEFINITION 1.1 If x is a construct, then x *exists mathematically* $=_{df}$ For some C, C is a set, class, or category, such that (i) x is in C, and (ii) C is specified by an exact and consistent theory.

(Of course the construct may have been conceived before being included in a set, class, or category described by a theory. But this psychological or historical circumstance is irrelevant to our concern, which is the mathematical legitimacy of the construct in question.)

What about the containing set, class, or category itself? In some cases it exists (formally) by virtue of being contained in a larger object, so we can answer the question 'Where (in which set, class, or category) is x?' But if the more comprehensive object happens to be the totality of constructs, then the question makes no more sense than the question 'Where is the physical universe?'.

By shifting the context from mathematics to some other field of knowledge, existence problems and existence statements (such as theorems) can be analyzed differently. For example, the formula "$(\exists x)Fx$", usually read 'There are F's', can be assigned, among others, the following interpretations:

(i) *ontological*: some real things possess the property represented by predicate F;

(ii) *psychological*: for some rational being y, and some time z, y thinks at time z that some objects possess property F;

(iii) *pragmatic* (or intuitionistic): some object(s) can be found or constructed with property F.

Clearly, none of these alternative readings is strictly mathematical. The ontological interpretation, legitimate in any factual discipline, calls for an empirical investigation to determine whether, in fact, there are material things possessing the property represented by the concept F. The second interpretation, legitimate in psychological contexts, stretches far beyond mathematics for making a claim of real existence; moreover, unlike the original statement, the psychological analysans is irrefutable. Finally, the pragmatic or intuitionistic interpretation – which is often adopted in the classroom and in computer science – produces a proposition that may not hold with the means available to the individual who adopts it. In any event all three interpretations go beyond the standard mathematical interpretation of the formula "$(\exists x)Fx$", which is simply this: *Some* constructs possess property F. (Equivalently: The set, class or category determined by F is nonempty.)

(By the way, what holds for the nonmathematical interpretations of "$(\exists x)Fx$" holds, *mutatis mutandis*, for the interpretation of any other mathematical formula. In doing or teaching mathematics we often indulge in extramathematical interpretations. Although such license can be heuristically or didactically effective, it can also be seriously misleading. For example, if we were to read the identity "$12 = 7 + 5$" as "12 can be decomposed or analyzed into 7 and 5", then it would cease to be identical with "$7 + 5 = 12$", and so the '$=$' sign involved in the two equations would not designate the identity relation, which is symmetrical. Likewise, if we interpret the fundamental theorem of algebra in pragmatic terms, as asserting that *we can find n* not necessarily different roots of a polynomial equation of degree n, then every time we fail to find such roots – as will often be the case for $n \geq 5$ – we may feel entitled to conclude that we have refuted the theorem. In short, within pure mathematics only mathematical interpretations are legitimate, all others are spurious. Nonmathematical interpretations of mathematical formulas are legitimate only in extramathematical contexts, such as physics and economics.)

The problem of mathematical existence is widely felt to arise only in relation with infinite sets. Thus intuitionists have no difficulty with finite sets but take extraordinary precautions – to the point of adopting a logic of their own – when it comes to infinite sets. And formalists concede that "Infinite totalities do not exist in any sense of the word [...] Nevertheless [...] we should act *as if* infinite totalities really existed" (Robinson 1965, p. 230). In our view all sets, whether finite or infinite, and indeed all constructs, pose the same philosophical problem and they exist in the same way, namely formally (or conceptually or ideally) – provided of course they are well defined. In other words, from an ontological point of view there is no difference among constructs: the modest number 0 and the entire "real" line exist in the same fashion, namely formally. None of them is physically real, and every one of them (except the totality of constructs) exists as long as it can be conceived by someone as being a component of some set, class, or category.

Of course conceivability, a necessary property of constructs, may pose problems even with regard to "small" mathematical objects. However, such problems are solved the minute we recall that, far from being stray and lawless, mathematical objects are systemic and lawful. An example will show why this is so. Consider an arbitrary 100 digit number n, and form the enormously greater number $n!$ (read 'n factorial'), i.e. the product of all the positive integers up to n. Multiply the result by $n + 1$, to obtain $(n + 1)!$. Since $(n + 1)! = (n + 1)n!$, $(n + 1)!/n! = n + 1$. A computer may handle the RHS but perhaps not the LHS of this identity in the case we are considering. Yet no mathematician doubts the (formal) existence of $n!$, or the quotient of the two. All three are well defined objects: they are defined by the laws (axioms, definitions and theorems) of number theory. And this is all that matters in mathematics: systemicity and lawfulness.

An obdurate empiricist might rejoin that he is prepared to admit large numbers, sets, etc., provided they are finite, for such constructs can somehow be imagined by extrapolation from smaller objects of the same type. But he refuses to countenance infinite numbers, sets, etc., because these are inaccessible to imagination, which (he believes) is always tied to sense experience (or to action in the case of pragmatism). In particular, the empiricist and the pragmatist may ask: How can you assert that there are infinitely many integers, if counting them all would take an infinite time, and would thus be impossible? (This question was posed by my daughter Silvia at age 5.)

The mathematician's answer to the above question is this. Although we cannot produce *every* single integer, *in principle* (though perhaps not in actual psychological fact) we can produce *any* integer by adding 1 to its predecessor. (Note the differences between "in principle" and "in actual fact", and between "every" and "any".) In general, we produce and "handle" infinities by using the finitary laws – e.g. of iteration or recursion – that they satisfy. For example, the sum of the powers of $\frac{1}{2}$ can be calculated as follows. Set $S = \frac{1}{2} + (\frac{1}{2})^2 + (\frac{1}{2})^3 + \ldots$ Multiply both sides by 2, to obtain $2S = 1 + \frac{1}{2} + (\frac{1}{2})^2 + (\frac{1}{2})^3 + \ldots = 1 + S$, whence $S = 1$.

In performing the preceding calculation we did not have to think of every member of S. We thought of the infinite series S as a whole defined by the law that generates it, a law that is finitary. (Actually we should have started by checking whether the sequence $\frac{1}{2}, \frac{1}{2} + (\frac{1}{2})^2, \frac{1}{2} + (\frac{1}{2})^2 + (\frac{1}{2})^3$, etc. converges to a limit.) We do not have to suppose, with the idealist philosophers, that the human mind has direct access to infinity. All we have to suppose is that the infinity in question is systemic and lawful. To put it in psychological terms: When faced with a (soluble) mathematical problem, whether or not it involves infinities, our task is to use or construct a finite number of functions pairing the inputs (data) to the outputs (solutions).

It might be thought that the preceding considerations hold only for specific sets, such as the real line and the complex plane. It might be doubted that our views hold also for abstract or nondescript sets, such as those occurring in the formula "$A \cup B = B \cup A$", where A and B are (or rather designate) sets. That is not so: if a set is abstract, i.e. nondescript or faceless, we handle it as a totality, assuming that it satisfies the algebra of classes. Systemicity and lawfulness legitimate our dealings with abstract sets, whether finite or infinite, denumerable or non-denumerable. The same holds, *mutatis mutandis*, for all the other mathematical objects.

In short, we must distinguish conceptual from material existence. Whereas to be a material thing is to be able to change in some respect, to be a construct is to be conceivable; and to be a mathematical construct is to be *conceivable consistently*. And, whereas material things are real, constructs are not; in particular, mathematical constructs are fictions. Consequently our materialistic ontology (Vols. 3 and 4), which refers to material things, does not hold for constructs. And our epistemology (Vols. 5 and 6) applies to the constructs purported to represent material things, but not to the rest: it is realistic with respect to the former, and fictionistic with regard to the constructs of logic and pure mathematics. (See Sect. 6 for the consistency of this dualism.)

2.2. *Mathematics and Reality*

How does formal (conceptual) existence relate to real (material, concrete) existence? We have already met one such relation, namely that between creators and creatures, i.e. inquiring animals and their conceptions. We have said that all constructs are created by higher animals: no thinking animal, no construct. (This is of course a denial of Platonism.) More precisely, we have characterized a construct roughly as an equivalence class of neural processes. This characterization accounts for the fact that, regardless of the ways in which different animals (or a single animal at different times) think of a given construct, they always think of the same construct despite their physiological differences.

(It might be objected that our characterization of a construct presupposes the notions of relation and of class. True, but this does not involve identifying constructs with brain processes. It is not that the processes themselves rely on such notions: they occur, and are similar in certain respects, whether we happen to know it or not. Likewise the members of a given chemical, biological or social species are similar to one another, and such similarities are objective, provided the species concept is well defined, even though every species is a set, and therefore a concept rather than a concrete thing. The notions of class and relation, as well as many other concepts, are required to account for reality but they are not themselves real. Our view raises interesting problems: see Torretti 1982.)

Another possible relation between formal and real existence is the one obtaining between a mathematical system and a physical (or biological or social) one. In other words: How is mathematics related to reality? This is of course a subproblem of the general question: What is the relation between ideas and the external world? If we look for an answer in the history of ideas we may be easily misled for, by conveniently disregarding counterexamples, we are likely to find cases confirming almost any of our philosophical prejudices. Thus from the fact that *some* mathematical ideas have originated in practical concerns, or have ended by being used in science or technology, one might be tempted to "conclude" that *every* mathematical object represents some trait of reality – the empiricist, pragmatist, and vulgar materialist thesis; or that every thing is identical with, or at least an imperfect copy or realization of, some mathematical object – the objective idealist thesis.

Both "conclusions" are wrong, as shown with the help of our theory of reference (Vol. 1, Ch. 2). Thus the laws of the algebra of classes – such as "$A \cup A = A$" – hold regardless of the nature of the members of the classes concerned. And the axioms governing the relation "less than" are independent of the nature of the elements of the set on which it is defined. Mathematics is *ontologically noncommittal*, and this is why it can be employed as a tool in constructing theories representing things of many different kinds – or none. Indeed the same mathematical systems ("structures") are likely to occur and recur in a great many different research fields, each time together with a different interpretation. Thus the real line can be used to represent a spatial direction, time, the degrees of a "quantity" such as energy, and much else. However, such interpretations (or semantic assumptions) are not part of pure mathematics: they are part of factual theories. (Recall Vol. 2, Ch. 6, Sect. 3.) The propositions in pure mathematics are about purely conceptual objects such as sets and functions.

If mathematics does not represent the world, if it is not the most general science of reality, then in particular it does not account for change. Shorter: mathematical objects are timeless. (But of course mathematical activity is temporal – like any other process.) Any mathematical description of real change involves some semantic assumptions whereby certain mathematical objects are assumed to represent nonmathematical objects such as places, times, velocities, or some other properties of real things. For example, the formula $\ulcorner y = \frac{1}{2}ax^2 \urcorner$ will represent, in one context, the distance travelled by a freely falling body (acceleration a) after time x; in another, the average power of a circuit (resistance a) through which alternating electric current of intensity x flows. Change occurs in the things represented by the mathematical constructs, not in the latter.

True, a moving body is sometimes said to integrate its equations of motion; a propagating wave, to integrate its field equations; a chemical reaction, to integrate its chemical kinetics equations; and the neurons in the visual cortex, to iterate certain Lie operators. But all these are metaphors. What is meant is that motion, chemical reaction, vision, or some other process, is *described* by certain mathematical operations. We do not *identify* processes with mathematical operations but assume that the latter can correctly *represent* the former. The representative exists conceptually (formally), the represented materially (really). The fact that in describing change we often employ the same mathematical objects, e.g. second order ordinary or partial differential equations, suggests that we muster only a somewhat limited conceptual repertoire. Nature does not care which concepts we employ to account for her. But we do care.

The contrast between immutable mathematical constructs and changing reality has baffled many thinkers for many centuries. Thus Plato believed that, since mathematical objects are immutable and perfect, and mathematical knowledge (unlike our knowledge of nature) is certain, it would never be possible to have a science of nature, which is always in flux and therefore imperfect. Only mathematics is science proper (*episteme*); everything else is opinion (*doxa*). "As being is to becoming, so is truth to belief" (*Timaeus* 28). Bergson (1907) too believed that intelligence, in particular science, is impotent to capture change and, in particular, novelty. And, just as Plato had stated that we can imagine myths about the world, but are incapable of building a science of nature, so Bergson claimed that intuition can grasp the mutability and creativity that elude intelligence.

More recently the same problem has been a philosophical motivation for the construction of an important new theory in the foundations of mathematics: topos theory. And some leading mathematicians have recently come to believe that set theory cannot be the correct foundation of mathematics because sets have an unchanging membership, and so they cannot adequately represent the mutability characteristic of reality, which, it is held, mathematics should strive to reflect. (See Lambek 1982.)

Our rejoinder to those who puzzle the contrast between the immutability of mathematical constructs and the mutability of reality is that the very nucleus of every advanced science is a set of equations of change of some sort of other, which equations involve unchangeable mathematical objects such as functions. For example Maxwell's equations, which do not propagate, serve to represent the propagation of electromagnetic waves; and the equations of population genetics serve to represent an aspect of the evolution of biopopulations. The reason for this ability of unchanging mathematical objects to represent change is, to repeat, that in factual science the mathematical objects concerned are taken together with semantic assumptions ("correspondence rules"). The latter specify what are the referents of those constructs and what properties of them they represent. Obviously, an examination of pure mathematics will not yield such semantic assumptions because pure mathematics is ontologically noncommittal. Only a semantic analysis of factual theories, e.g. in chemistry or in biology, will yield them. (See Vols. 1 and 2 for the concept of a semantic assumption and examples.)

To understand how scientists manage to account for change with the help of unchanging mathematical constructs, it may help to consider a few instances. *Example 1* Call L Archimedes' law of the lever, an algebraic expression involving symbols interpreted in mathematics as numbers, and in physics as values of distances and weights. (Recall Sect. 1.2.) The condition of nonequilibrium or movement of such a system is obviously the negation of L. But note that *not-L* does not involve the notion of time. *Example 2* Consider a continuous function f from reals to reals: no change so far. The function can be made to represent a continuously varying property of some real thing if $\mathbb{R}$ is made to represent time, and the value $f(t)$ of f, for t in $\mathbb{R}$, the value of the property concerned at time t. Think of the elongation $\sin wt$ of a vibrating mass point.

Example 3 To represent a collection of changing membership, such as the population of Paris, we can introduce the notion of a variable (or time dependent) collection (Vol. 3, Ch. 5, Sect. 2.5). A variable (or time dependent) collection S_t is definable as the set of all the individuals possessing the given (defining) property P at a given time t, i.e. $S_t = \{x|Pxt\}$, where $t \in \mathbb{R}$ is interpreted as an instant. (Think of Pxt as "person x lives in Paris at time t".) The notion of an invariable collection can now be obtained by taking the union of all the S_t over time, i.e. $S = \bigcup_{t \in \mathbb{R}} S_t$. And if necessary we can define a time dependent membership relation $\in_t$ through the condition:

$$b \in_t S =_{df} b \in S_t.$$

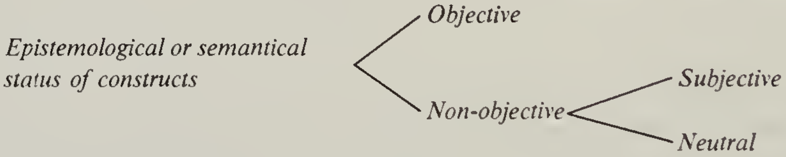
Thus to say that Aristotle *is* human, or belongs to the human species S (invariable membership), amounts to saying that Aristotle belong*ed* to the collection of humans who *were* alive at the time of Alexander. In short, the notion of a variable collection is definable in terms of the mathematical concept of a set and the ontological (or protophysical) concept of time. (However, it is also possible to adopt the notion of variable collection as undefined.)

The preceding examples illustrate our thesis that, although pure mathematics is ontologically noncommittal, it provides a conceptual framework which, when enriched with the suitable semantic assumptions, can be made to account for our restless universe. In the last two cases the (or rather a) concept of time, which is alien to pure mathematics, occurs essentially, in Example 2 as the independent variable, in Example 3 as an index. It was taken as a primitive concept to be elucidated elsewhere, namely in ontology or in protophysics. However, any exactification of any concept of time, or of any other factual concept for that matter, is bound to employ some mathematical (and also some nonmathematical) concepts. (See Vol. 3, Ch. 6, Sect. 3.). Hence the notion of a variable collection, defined in Example 3, cannot occur in the foundations of mathematics except as an undefined concept. Shorter: We need the a priori science of mathematics to build the a posteriori sciences of the real world. (More on this in Sect. 4.1.)

Although mathematics is not semantically objective, for it does not represent the external world, it is undoubtedly objective in some other sense, in which art and religion are not. Thus we all agree that the rational numbers are denumerable, that the rotations about a fixed point constitute a group, and that some differential equations can be integrated by the method of the Laplace transform. Moreover such agreement is not merely a matter of fashion or arbitrary convention: it is a result of reasoning. How can that be? The answer is simply that, once certain assumptions have been adopted, we are rationally committed to admitting their logical consequences. Moreover such assumptions are not stray but form systems (theories). So, the objectivity of mathematics consists in the systemicity and lawfulness of its objects, not in that mathematics is a sort of universal physics. Shorter: mathematics is *methodologically objective*, not semantically or epistemologically objective.

Does our denial that mathematics is semantically or epistemologically objective entail that it is subjective? Certain formulations of mathematical statements do cause this impression. For example, Zermelo's principle is sometimes stated "Every set *can be* well ordered", and Euclid's 5th postulate is often stated "Through any point outside a given line *it is always possible to draw* a single parallel". However, these formulations are not strictly mathematical: they are pragmatic interpretations of "*There is* a well ordered set containing the elements of any given set", and "*There is* a single parallel through any point outside a given straight line". (In both cases 'there is' designates the concept of conceptual existence.) None of these statements refers to subjective experience or to action: they only assert the formal existence of certain mathematical constructs.

How is it possible for something, in particular mathematics, to be *neither* objective *nor* subjective in an epistemological or a semantical sense? The reason is that *objective/subjective* is not a dichotomy, for *tertium datur*, namely epistemological (or semantical) *neutrality*. Indeed, the correct dichotomy is objective/non-objective, where "non-objective" equals the exclusive disjunction of subjective and neutral. I.e.,



From an epistemological (or semantical) viewpoint mathematics is neither subjective (intuitionism) nor objective (Platonism and dialectical materialism) but *neutral*, because it is neither about subjective experience nor about an autonomously existing world. Of course mathematical creation presupposes the existence of creators, i.e. living mathematicians working in a favorable culture. (No mathematics without mathematicians: mathematics will end with the last mathematician.) But the point is that, although mathematics is created by real beings, it is neither independently real nor, by itself, representative of reality. Yet nothing prevents us from *pretending* that mathematical objects exist in a *sui generis* fashion (i.e. formally), and everything encourages us to use mathematics in our study of reality.

Mathematical objects are then ontologically on a par with artistic and mytholological creations: they are all *fictions*. The real number system and the triangle inequality axiom do not exist really any more than Don Quijote or Donald Duck. Nor is the difference epistemological, for some mathematical constructs, just as some artistic and mythological fictions, are idealizations of real things or features of real things. Nor, finally, is mathematics distinguished from art and myth by its certainty, even though the yearning for final certainty has often been a powerful motivation for mathematical research, particularly in the foundations of mathematics. But the successive crises in the latter, the current multiplicity of inequivalent set theories, and the continuing controversies over the axiom of choice, have taught us that final certainty is a mirage. True, mathematics is more certain than factual science, which is in turn more certain than other research domains. Still, mathematical research is not error-free. True, mathematics has foundations, but the latter are not unique and they are not immutable.

The crucial differences between mathematical fictions and all others are found elsewhere. Unlike other fictions,

(i) mathematical objects, though devoid of factual reference, are not totally free inventions, let alone lies or products of self-deception: they are constrained by laws (axioms, definitions, theorems); consequently they cannot possibly behave "out of character" – e.g. there can be no such thing as a right angle equilateral triangle, whereas even mad Don Quijote is occasionally lucid, and even gentle Mickey Mouse is occasionally cruel;

(ii) mathematical objects exist (formally) either by postulate or by proof, never by arbitrary fiat;

(iii) mathematical objects are theories or referents of theories, whether full-fledged or in the making, whereas myths, fables, stories, poems, sonatas and paintings are nontheoretical;

(iv) mathematical objects and theories are fully rational, not intuitive, let alone irrational (though of course mathematical intuition is acquired with practice);

(v) mathematical statements must be justified in a rational manner, not by intuition, revelation, or experience;

(vi) far from being dogmas, mathematical theories are based on hypotheses that are given up if shown to lead to contradiction, triviality, or redundancy;

(vii) mathematical theories are linked together forming a supersystem; thus logic employs algebraic methods, and number theory resorts to analysis; on the other hand there is no such thing as a coherent system of artistic or mythological creations;

(viii) mathematics is neither subjective nor objective, but ontologically noncommittal; only the process of mathematical invention is subjective, and only living mathematicians are real;

(ix) mathematical objects and theories find application in science, technology, and the humanities;

(x) mathematical objects and theories are socially neutral, whereas myth and art often support or undermine the powers that be.

So much for the general traits of mathematics. Let us now take a somewhat closer look at our science, starting with logic, the most basic branch of mathematics and the intersection of mathematics and philosophy.

3. LOGIC

3.1. *Logic Lato Sensu*

Logic originated as the art of conversation and, in particular, argument. Eventually it developed into the theory of deduction. We shall call this restricted construal of 'logic' *logic stricto sensu* or *basic logic*. This is the way philosophers usually conceive of it. Mathematicians nowadays include in logic not only the theory of deduction but also model theory (not to be mistaken for modal logic), set theory, recursion theory, and often also category theory. (See e.g. Barwise Ed., 1977.) We shall call this *logic lato sensu*. There is an even more comprehensive construal of "logic": some philosophers have talked about 'the logical structure of the world', 'the logic of discovery', 'the logic of decision making', 'the logic of verbs', and even 'the logic of the situation'. We shall disregard these metaphorical uses of the word 'logic'.

In this section we shall tackle a random sample of philosophical problems about logic *lato sensu*. In particular we shall ask: What are predicates and propositions?, How is $\exists$ to be interpreted?, What is a model of an abstract theory?, What is the philosophical import of Gödel's incompleteness theorem?, What is all the fuss about the axiom of choice?, Are sets necessarily the building blocks of mathematics?, and What is logic?

To start with the first of the problems in our list: as we saw in Vol. 5 (Ch. 5, Sect. 2.1), *predication* or attribution is a fundamental cognitive operation. In ordinary knowledge, factual science, and technology, predication consists in conceiving predicates that are hoped to match the properties or relations observed or conjectured to hold in reality. Thus the conceptualization of ordinary experience with ponderables results in predicates such as "heavy" and "light", "heavier than" and "lighter than", "weight of a body" and "weight of a body relative to a reference frame".

Predication is sometimes sparked off by experience but it can go far beyond the latter, and indeed in two ways. Firstly, in all walks of life we employ transempirical predicates such as "weightless", "unseen", and "shoddy or expensive". Modern science and technology teem with such predicates: think of "state (of a thing)", "ionization", and "balance of payments". Secondly, mathematicians keep inventing predicates that do not correspond to any substantial properties, i.e. properties possessed by material things. Think of "predicable", "is a logical consequence of", "definition", and "consistent" – to mention only a few logical and metalogical predicates.

This brings us to an important yet generally overlooked difference between the predicates occurring in mathematics and all the others. Whereas a mathematical predicate *is* (identical with) a property of a construct, a predicate in any other field of knowledge is (rightly or wrongly) supposed to *represent* a property of some substantial (material, concrete, real) thing. (We drew this difference in Vol. 3, Ch. 2, Sect. 1.) Thus it is indifferent whether we call 'differentiable' (with reference to functions) a predicate or a property. On the other hand we must draw a difference between a property of a material thing and any of the predicates that may represent that property. First, because predicates are constructs whereas substantial properties are features of real things, hence inseparable from the latter. (Real things "come" with all their properties.) Second, because one and the same property can be conceptualized differently in different theories: think e.g. of the mass and length properties in classical and in relativistic mechanics. (Incidentally, the fact that different theories may choose different attributes to represent one and the same substantial property confirms critical realism as formulated in Bunge, 1983b. And it refutes idealism – which identifies attributes and properties – as well as naive realism, which postulates a 1:1 correspondence between attributes and properties.)

The most basic, and therefore the coarsest, formalization and systematization of predication (or attribution) is the predicate calculus. This is logic *stricto sensu*: it studies predicates and the formulas and deductive inference patterns in which they occur. (Examples of kinds of predicate: unary such as "heavy", binary such as "heavier than", and in general n-ary. Example of a formula: open such as "Axy", and closed such as "$(\forall x)(\exists y) Axy$". Example of a rule of inference: $Aa \vdash (\exists x)Ax$.) In turn, the predicate calculus is a refinement of the propositional calculus, which does not analyze the basic or atomic propositions, such as "a is heavier than b". Still, the predicate calculus is extremely coarse in comparison with the general theory of functions. Thus the latter allows us to analyze formulas such as "The telephone number of x is y", and "The production of x in dollars, during the year y, was z". Morever, unlike the predicate calculus, the theory of functions allows one to distinguish between different functions with the same arguments, such as the linear function, the sine function, and the logarithmic function. Logic jumbles all of the infinitely many functions of one real variable into the binary predicate.

The notions of a predicate and of a proposition are, of course, elucidated by the predicate calculus. But a somewhat deeper study of those notions calls for the mathematical concept of a function (Vol. 1, Ch. 1, Sect. 1). Indeed, an arbitrary unary predicate P can be construed as a function (or many-one relation) from a domain A of objects to the set S of all the propositions involving P. In symbols, $P:A \to S$. The value of P at a in A is Pa, a member of S. Likewise, an arbitrary binary predicate R can be construed as a function from ordered pairs of individuals – i.e. from elements of the cartesian product $A \times B$ of two sets – to the set T of all the propositions involving R. In symbols, $R: A \times B \to T$. The value of R at $\langle a, b\rangle$ in $A \times B$ is Rab, a member of T.

For example, the predicate "predicable" can be meaningfully (though perhaps falsely) predicated of concepts, not of material objects. In other words, the domain of that predicate is the set of concepts. The fact that most concepts cannot be truly predicated of themselves (e.g. "even" is not even) is beside the point. The notion of truth does not occur in our analysis of a predicate. (It does occur in Frege's queer notion of a predicate as a function from individuals to truth values: see Vol. 1, Ch. 1, Sect. 1.3.)

We have clarified the functional relation between predicates (or attributes) and propositions without saying explicitly what these constructs are. Yet one answer to this question is tacitly included in the above: A proposition is a value of a predicate, and predicates are defined (implicitly and in broad outline) by the axioms of the predicate calculus. Still, this definition is psychologically and practically as unsatisfactory as the (correct) characterization of an electron as that which satisfies Dirac's theory of electrons. We would like to have an independent characterization of a proposition (and of an electron), not only to satisfy our intuition but also to be able to recognize it as such when meeting it, as well as to be able to identify propositions across the various alternative logical theories (such as the classical and the intuitionistic predicate calculi).

One possible move is to say that a proposition is an object composed of concepts, in particular predicates. This is true but insufficient, for constructs can be combined in ways that do not yield propositions. (E.g., 'Dogs are' is a meaningless combination of meaningful words.) Something more is needed, namely a particular structure: the concepts have to be put together in a proper way. In other words, we may conceive of a proposition as a (conceptual) system composed of concepts and having a structure of a certain kind. Propositions with the same composition but different structure, such as "Dogs are animals" and "Animals are dogs", may be called *isomeric*. On the other hand propositions with the same structure but different compositions, such as "Space is divisible" and "God is omnipotent", may be called *isomorphic*. However, none of this is too helpful because it does not specify the right kind of structure, i.e. the one that combines concepts into propositions. (Saying that the structure is specified by the formation rules of the predicate calculus is quite true but not to the point, since our goal was to characterize the notion of a proposition independently of that theory.)

A second move is to resort to semantical considerations. Thus we may say that predicates are objects with a sense (or content) and a reference. True but insufficient, because in our theory propositions too have both properties. Besides, our own theories of sense and reference (Vol. 1) presuppose that we know what a predicate and a proposition are. Another possible way out is to switch from considerations of meaning to considerations of truth. As a matter of fact the most popular "definition" of a proposition is this: A proposition is whatever has a truth value. However, this characterization has two defects. One is that it, too, presupposes some knowledge of what a proposition is, at least enough to distinguish the semantical concept of truth from the concepts occurring in the phrases 'a true gem' and 'a true friend'. The second shortcoming is that the great majority of the propositions occurring outside pure mathematics are at best approximately true, at worst undecidable.

The balance of our exercise is this. First, it is true that propositions are systems of concepts, and that the structure of such systems is specified by the rules of formation of the predicate calculus. (E.g., if M and N are predicates of rank m and n respectively, and O is an arbitrary set of objects, then $Mab\ldots m$, $Nde\ldots n$, $\neg Mab\ldots m$, and $Mab\ldots m \vee Nde\ldots$ with $a, b, \ldots, m, n \in O$, are propositions.) Likewise it is true that both predicates and propositions have senses and references. Finally, it is also true that some propositions have a truth value, though not necessarily 0 or 1. However, our semantical considerations do not serve to define either predicates or propositions. What they do is, first, to *supplement* the syntactic characterization given by the predicate calculus. Second, they serve as indicators or *criteria* of predicateness or propositionhood (pardon the neologisms). For example, if P is a construct that combines objects a and b into a construct of the form Pab, then P is a predicate of rank 2, and Pab a proposition. (Incidentally, the fact that such criteria cannot double as definitions refutes the constructivist or operationist doctrine of definition. Likewise, convergence criteria or tests presuppose rather than replace the definition of convergence.)

What happened to our desideratum of being able to identify a given proposition across all possible logical calculi? After all, (*a*) the point of building different theories of objects of a given kind is to attempt to capture in the best possible way the essential properties of such objects; but (*b*) if two constructs are defined (implicitly) by different theories (not just different presentations, e.g. axiomatizations, of one and the same theory), then they are not exactly the same constructs. (Recall from Sect. 1, 2 that constructs are characterized by their sense and reference, so that even if two theories refer to the same objects, they can have different senses, hence they describe different constructs.)

Note the difference between this problem and that of identifying the referents of two different factual theories, such as two alternative theories of the electron. In this case there is a single real object modeled now in one way, now in another; i.e., there are two different, possibly rival, concepts of the electron. But both constructs, if they are to command the attention of scientists, will have to satisfy the description of an electron as an entity with such and such mass, spin, and electric charge values. In the case of propositions we cannot resort to such empirical characterizations because propositions are *êtres de raison*, not material entities that can be manipulated in the laboratory. In this case our considerations must be intraconceptual and, more precisely, metatheoretical.

What we may do in this case is the following. We may stipulate that two theories deal with propositions (and also possibly with predicates) if, and only if, (*a*) they define (either implicitly or explicitly) operations such as "or", and (*b*) they share some tautologies, such as "If p and q then q", or some rules of inference, such as the *modus ponens*. (Caution: do not require that the principle of non-contradiction be included among the shared tautologies, for it cannot even be stated in negationless logics, and it is not assumed in the so-called paraconsistent logics. More on this in Sect. 3.2.)

The second item in our agenda concerns the proper interpretation of the *"existential" quantifier* ∃, which we prefer to call *particularizer* or *indefinite* *quantifier*, to distinguish it from both the *universalizer* $\forall$ and the *individualizer* (or descriptor) ι. As it is well known, most logicians believe that "Existence is what existential quantification expresses" (Quine, 1969, p. 97). We have argued in Sect. 1.3 that this is a mistake: that there are two radically different concepts of existence, which $\exists$ does not distinguish: formal (or conceptual) and material (or real) existence. We have also argued (Vol. 3, Ch. 3, Sect. 4.3) that, far from exactifying either of these concepts, $\exists$ exactifies the concept of *some*. Thus, "$(\exists x)x$ is a ghost" is not to be read as "There are ghosts" but as "Some individuals are ghosts" – namely some figments of our imagination. (An alternative interpretation of "$(\exists x)Gx$" is: The formula "Gx" is satisfiable. Equivalently: Some substitution instances of "Gx" are true". Though different, the two interpretations – which occur in *Principia Mathematica* – are equivalent: Something has G if some substitution instances of "Gx" are true and conversely.)

To clarify the above consider the statements

$$\text{There are wicked ghosts.} \tag{1}$$

$$\text{Some ghosts are wicked.} \tag{2}$$

According to Russell, Quine and most other logicians these two statements are identical, and they (or rather it) must be formalized as "$(\exists x)(Gx \;\&\; Wx)$". In our view (1) and (2) are different: the first has an existential import, the second does not. Consequently the first step towards an adequate formalization of (1) is to rope in the existence predicate E_M defined in Vol. 3, Ch. 3, Sect. 4.3. Being a predicate, E_M operates on the individual variable x like any other predicate. The result is: $E_M x \;\&\; Gx \;\&\; Wx$. If '$M$' is interpreted as mythology, our formula reads: In mythology there are individuals both ghostly and wicked. If we wish to obtain a proposition (closed formula) we prefix the particularizer $\exists$, to obtain $(\exists x)(E_M x \;\&\; Gx \;\&\; Wx)$, read "Some of the ghosts (that exist) in mythology are wicked". Only now have we made an ontological commitment, whereas (2) remains ontologically neutral. (The commitment is serious if we presuppose that mythology is true, phoney if we do not.)

The preceding considerations have some interesting consequences. One concerns mathematical notation whenever existence statements are made. For example, the postulate of group theory asserting the existence of inverses is: For all x in G there is a y in G such that $x \circ y = e$ (where e is the unit or identity element of G). This axiom is usually written in the form

$$(\forall_G x)(\exists_G y)(x \circ y = e). \tag{3}$$

But, if our interpretation of $\exists$ is correct, the axiom should be rewritten as

$$(\forall_G x)\,(\exists_G y)E_G y\;(x \circ y = e). \tag{4}$$

All the other existence statements should be symbolized in a similar fashion.

A second consequence of our deontologization of the "existential" quantifier is that it disposes of Quine's objection to second order logic (which quantifies predicates) for its alleged "excessive ontological commitment".

A third consequence of our consideration is far more substantial and of a technical nature. The usual existential interpretation of $\exists$ is responsible for the abandonment of an important intuitive principle of ancient and medieval logic. This is the principle of *subalternation*, according to which what holds for all holds for some. (More precisely, "All Fs are Gs" entails "Some Fs are Gs".) This principle does not hold if "some" is replaced with "there exist". Thus "All ghosts are playful" does not entail that *there are* playful ghosts. But it does entail that *some* ghosts are playful – an assertion that, like its premise, has no existential import. If the existential interpretation of $\exists$ is avoided, the principle of subalternation can be allowed to stay.

Another technical consequence of our reinterpretation of $\exists$ is that the axiom "$Fa \vdash (\exists x)Fx$" stays too, but reinterpreted as "If an individual has a certain property, then *some* individual(s) have the same property". If we wish to conclude existence then we must assume it to begin with, reformulating the preceding principle as follows:

For any inhabited context C, $Fa \,\&\, E_C a \vdash (\exists x)\,(E_C x \,\&\, Fx)$.

Finally, our reinterpretation of $\exists$ has an important philosophical consequence concerning the nature of elementary logic. We can now confidently assert, *pace* Quine, that *elementary logic is ontologically neutral rather than committed.* The deceiving appearance of ontological commitment comes from the mistaken reading of $\exists$ as "there exists" instead of "some". Existence is not just a matter of quantifiers but of either postulation or proof. (For additional arguments in favor of the ontological neutrality of logic see Bunge 1974c.) So much for "some" and its confusion with "there is".

A related problem is that of *analyticity*. An analytic formula is one which is true by virtue of its form or of the meaning of the concepts it contains. In the first case we call it a tautology, in the second a tautonymy. Examples: "All bachelors are bachelors" and "All bachelors are unmarried men" respectively. Quine (1952) challenged the analytic/synthetic dichotomy on the ground that we have no satisfactory concept of meaning, hence of synonymy, hence of tautonymy. But we claim to have supplied such a theory in Volumes 1 and 2 of our *Treatise*. Hence we keep the analytic/synthetic distinction although it is far less important than the formal/factual one. But we do admit that analyticity is *contextual* for both tautologies and tautonymies. The former is clear given the multiplicity of logics. Thus the excluded middle principle is tautologous in classical logic but not in intuitionistic logic. As for tautonymies, consider the following example. In the context of ordinary knowledge "Anything that bounces is solid and resilient" is tautonymous by virtue of the meaning of "bounces". Not so in the context of atomic physics, where elementary "particles" can bounce off atomic nuclei. Moral: To be correct, logical analysis must take propositions in context; grammatical analysis is parallel.

Let us now say a few words about *model theory*. But first let us repeat the warning issued in Vol. 5, Ch. 9, Sect. 1.2 concerning the ambiguity of the term 'model'. In factual science and technology a model is a specific theory supposed to represent (poorly or accurately) natural, social or artificial things of some kind. Thus one speaks of a mathematical model of a molecule, or a neural system, or an irrigation network. We call this the *epistemological* concept of a model.

In logic, on the other hand, a model is an example or "realization" of an abstract system, and a theory of such a model is an interpretation of the corresponding abstract theory. For example, there are arithmetical and geometrical models (examples) of abstract sets. And, in turn, abstract sets (along with the integers, matrices, and other constructs) are models of a ring, i.e. they satisfy all the formulas of ring theory. The key word here is 'satisfaction': a model of an abstract system ("structure") *satisfies exactly* all the conditions that define the system. The counterpart of 'satisfaction' in the case of scientific and technological models is 'representation': these models *represent approximately* their referents. A model of a lattice is another lattice: both are conceptual systems. On the other hand a mathematical model of a material entity is not another chunk of matter but a specific theory. The failure to distinguish the logical from the epistemological concepts of model (as in Sneed 1971 and Stegmüller 1976) has given rise to an utterly artificial philosophy of science. (See Truesdell 1984 for proof that this philosophy is remote from real science.)

Model theory deals exclusively with models in the logical sense, and has hardly anything to say about models in the epistemological sense. Two characteristic problems in model theory are to determine the relation between an abstract system and its models, and to characterize the relation between the latter. The reasons that model theory is nowadays included in logic are that (*a*) it has succeeded in elucidating one of the two notions of logical consequence, namely that of semantic entailment ($\models$), and (*b*) it has shown that an abstract theory can be proved consistent by exhibiting a model of it, i.e. by translating it into a "concrete" (but still purely conceptual) theory.

Model theory is of interest to philosophy for a number of reasons. Firstly, because it constitutes a natural enlargement of elementary (first order) logic. Secondly, because it exhibits the limitations of the purely syntactic approach to logic (which defines "⊢" but not "⊨"); and, with it, the impossibility of nominalism, which discards meaning and truth. Thirdly, because model theory highlights the difference between two kinds of mathematical system and, correspondingly, of mathematical theory: abstract (or "formal") and "concrete" (or specific). Thus an abstract group is one the nature of the elements of which is not specified beyond the axioms of general group theory. On the other hand the group of all rotations of space about a fixed point is a model of a group – hence another group – and it can be studied independently in geometry. Such rotations exemplify ("realize", as it used to be said) the abstract concept of group operation, and the theory of such rotations is a theory of that model of a group. (This is of course the gist of Klein's Erlangen Program: to ferret out the group-theoretic structure underlying every geometry.) An alternative model or example of a group is the set of integers with addition as the group operation, subtraction as the inverse operation, and zero as the unit element. Abstract theories are presumably the kind of mathematics Russell (1901) had in mind when he wrote that "mathematics may be defined as the subject in which we never know what we are talking about, nor whether what we are saying is true". This formalist description of mathematics fits "formal" (i.e. abstract) theories, not their models.

There are further reasons for philosophers to become interested in model theory. One is that it has contributed to elucidating the notion of formal truth in contradistinction to that of factual truth. Indeed, the notion of satisfaction of an abstract formula in a model exactifies the coherence theory of truth, according to which (formal) truth consists in the matching of conceptual systems. (Recall Vol. 2, Ch. 8, Sect. 2.1, and note that this notion of truth is radically different from that of adequacy.) A second reason is that model theory has clarified the notion of translation or interpretation of one theory into another – as when the formulas of group theory are translated into number-theoretic or geometric formulas. A third reason for a philosopher to become interested in model theory has to do with the implicit definition of the basic (primitive) concepts of a theory. Since any consistent abstract theory has an unlimited number of models, the postulates of the theory can define only the "essential meaning" of its primitives. (Equivalently: the "essence" of a family of models is the underlying abstract axiomatic foundation, which sketches the structure common to all models.)

A fourth reason is that, by giving birth to non-standard analysis (a generalization of the infinitesimal calculus), model theory has proved that logic *lato sensu* is not only an analytic tool but also a constructive one, and in any case anything but an idle formal game. And a fifth reason is that model theory has contributed to making mathematics more systemic or cohesive, by showing that certain previously disconnected theories are nothing but different "realizations" (interpretations or translations) of one and the same abstract theory – or, to indulge in metaphor, different embodiments of a single structure.

It is now the turn of the most famous (and misunderstood) of all metamathematical results, namely *Gödel's two incompleteness theorems*. The first asserts, roughly, the existence of some undecidable formulas in every consistent theory containing arithmetic – e.g. in arithmetic itself. (Formula A is said to be *undecidable* in theory T if neither A nor $\neg A$ is a theorem of T.) The second theorem states that no theory of that type can be proved consistent with the sole resources of the theory itself. (See e.g. Stoll 1963.) The first incompleteness theorem entails that formalization does not capture all our intuitions about natural numbers; the second, that mathematics as a whole cannot be proved to be consistent. (See Wang, 1974 for these and other implications.)

These results have made no impact on the practice of the working mathematicians, who carry on business as usual: they only "still produce nightmares among the infirm" (Smorynski 1977 p. 825). On the other hand they shattered Hilbert's noble program of proving by finitist methods the consistency of all mathematical theories. Consequently they had a profound impact on the foundations of mathematics, for which the year 1931 has been a watershed. Among the many remarkable results that followed was this: only a few theories are decidable (Tarski *et al.* 1968).

The incompleteness theorem made a strong impression on philosophers, particularly when misunderstood. The irrationalists read them as proof that reason had finally reached the end of its tether. And they likened the theorems to Heisenberg's "uncertainty relations", also misunderstood as setting a limit on human knowledge. Of course, mathematicians did not take a pessimistic view of the scope of human reason. If anything they admired Gödel for giving proof of a superior intellect: for suspecting, and then proving, totally unexpected theorems. And mathematics has made more advances in the half century that followed 1931 than in any other half century of its long history.

To a mathematician, Gödel's incompleteness theorems show only the limitations of axiomatics and, in particular, of formalization (a particularly rigorous and exhaustive form of axiomatics, where all the cards are put on the table). The theorems boil down to the assertion that, given a body K of mathematical knowledge including arithmetic, and a formal system F built with the aim of formalizing K, there will always be formulas in K that are neither provable nor refutable in F. This technical result is hardly of philosophical interest: it only reminds us, once again, that there are few perfect human creations.

What does have some interest is Gödel's claim (in Wang 1974 p. 9) that his theorems confirm his own Platonistic philosophy of mathematics, according to which the natural numbers exist objectively just like the stars. This claim would hold only if there were some *evidence* for the existence of Plato's Realm of Ideas. In this case one could place the informal bodies of mathematical knowledge (the Ks of the preceding paragraph) in that ideal realm, and would regard their formalization (the Fs of a while ago) as poor man-made copies of those allegedly self-existing eternal verities.

But there is no such evidence. On the contrary, the history of mathematics teaches that every body of mathematical knowledge has been created by some living brain. Resorting to Plato's Realm of Ideas in themselves is like invoking divine providence every time we experience a fact which we are unable to explain. Moreover, a philosophy about the nature of constructs cannot be self-contained but must be part and parcel of a comprehensive ontology. It must account not only for the peculiarities of the fictions which we decide to endow with (formal) existence, but also for the human activity that creates such fictions; i.e. that ontology must include a philosophy of mind. Obviously, mathematical Platonism falls short of these requirements and, in particular, it is inconsistent with the scientific view that there are no brain-free ideas. We shall come back to this matter in Sect. 5.2.

The next item in our agenda is the *axiom of choice*, which has become a bone of contention of such magnitude that it has divided the mathematical profession into those who use it, or at least do not object to its use, and those who reject it. Yet the axiom looks innocent enough: For every family F of pairwise disjoint nonempty sets, there exists a set C containing exactly one element in common with every element of F. (See Figure 1.2.) The main reason, or unreason, for the controversy is that the axiom is nonconstructive: it asserts the existence of an object, namely C, without showing how to construct it, i.e. without specifying the choice function that maps F into C. (Hence, except for their heterogeneous origin the elements of the set C are quite faceless, ergo indistinguishable.) Being nonconstructive, the axiom

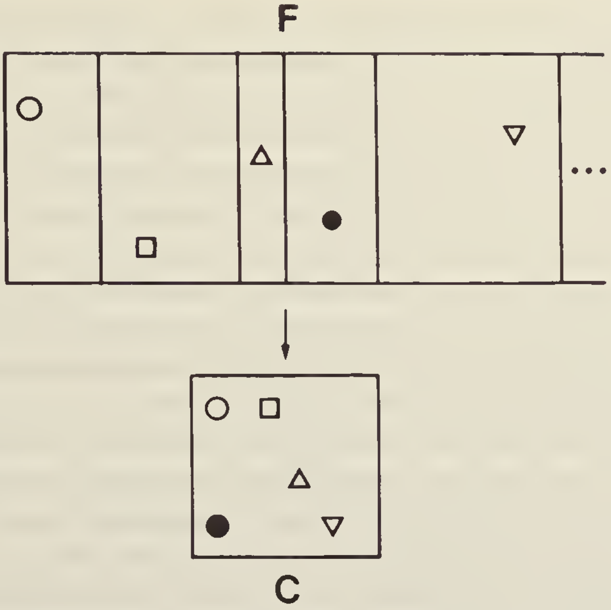


Fig. 1.2. The axiom of choice. Given an arbitrary partition of a set *F* into pairwise disjoint subsets, there exists a set *C* containing one representative of each of the subsets of *F*. No rule is given for forming *C* out of *F*.

of choice cannot be accepted by the constructivists, who abide by the maxim that only that exists which can be effectively constructed. (See Sierpiński 1958 Ch. VI for a large collection of interesting opinions on this matter, and Moore 1982 for an exhaustive history.)

For better or for worse, the axiom of choice is the only nonconstructive postulate in classical set theory: all the other postulates can be interpreted pragmatically as instructions for forming new sets from given sets. Moreover, it has been proved (1963) that, though consistent with the remaining axioms of set theory, the axiom of choice is independent of them. This makes it possible to build much of mathematics with the help of a set theory not including the axiom of choice.

On top of being nonconstructive, the axiom of choice gives rise to some "paradoxes", though not to contradictions. The best known of them is perhaps the Banach-Tarski paradox which states that, using the axiom of choice, one can cut a ball into a number of pieces that can be put together so as to obtain two balls of the same size as the original ball. (See Jech 1973.) However, as Fraenkel (1966) has stated, dispensing with the axiom leads to even stranger theorems. And, above all, it results in a considerable impoverishment of mathematics, for the axiom appears in dozens of different though equivalent guises and disguises – so much so that it has been used repeatedly by some of its critics (Moore 1982). After all, intuitability is a subjective property and therefore not a reliable indicator of truth or validity. We shall come back to this point in Sect. 5.1: for the time being suffice it to stress that the controversy over the axiom of choice, still raging since its inception at the beginning of our century, shows that mathematics is not free from uncertainty and is not dominated by consensus over fundamentals. We shall come back to these matters in Sect. 5.1.

Our next question is whether set theory is necessary and sufficient to build the rest of mathematics. Until recently all mathematicians, except the intuitionists, answered this question in the affirmative, and nearly all philosophers followed suit. (See however a warning note in Bunge 1962a.) Indeed, all of "classical" (non-intuitionistic) mathematics seemed to be built with the sole help of set theory and its underlying logic, namely the ordinary predicate calculus. The Bourbaki treatise culminated the process of reduction of all mathematical concepts to those of set and membership. It is therefore understandable that the attention of most workers in the foundations and philosophy of mathematics should have focused on set theory, generally regarded as the bedrock of mathematics.

This situation was altered radically in the 1960s by two unexpected events. One of them was the proof that there is no such thing as *the* theory of sets: instead, there is whole family of different set theories. (A while ago we saw one source of multiplicity: the possibility of accepting or rejecting the axiom of choice. There are other axioms, in particular the continuum hypothesis, that are also independent of the remaining axioms for sets, so they can be kept or dropped.) In any event a new question of method or policy seemed to emerge: When stating that the whole of mathematics is reducible to set theory, which set theory is meant?

The second event that changed the foundational landscape in the 1960s was the discovery (Lawvere 1966) that the concepts of set and membership, which are basic (undefined, primitive) in set theory, are definable in another theory, namely *category theory*, founded in 1945. (See MacLane, 1971.) The basic notions of this theory are those of morphism – an extension of the concept of a function – and morphism composition. This being so, one must own that the question at the end of the last paragraph was ill-conceived, for it presupposes that some set theory has got to be the right and ultimate foundation of mathematics. This presupposition has been shown mistaken: sets have turned out to constitute just a model of a category; i.e. categories are more abstract, fundamental and inclusive mathematical objects than sets.

Category theory deserves to be closely examined by the philosopher of mathematics for the following reasons. First, it is currently the most basic theory on top of first order logic: it is the contemporary foundation of mathematics – mark the qualifier ‘contemporary’. (See Figure 1.3.) Second, unlike all the other branches of mathematics, category theory deals wholesale with mathematical systems (“structures”), such as the totality of topological spaces. (These totalities are categories, not sets. And the way to deal with such totalities is by means of morphisms, some of which are structure-preserving whereas others are not. For example, a group morphism relates groups, whereas a certain “forgetful functor” relates groups to semigroups by “forgetting” the inversion operation.) Third and consequently, category theory has not only enriched mathematics but has also exhibited a previously hidden network binding together the various mathematical systems (“structures”) and consequently their corresponding theories. It is just as universal and cross-disciplinary as first order logic. Thus category theory is accomplishing one of the ideals of the founders of set theory, namely the unification of mathematics on the basis of the most general concepts.

After having sampled the problematics of the philosophy of logic, we are ready to face the last question in our agenda: *What is logic*? (See Kneale and Kneale 1962 for the history of this question, and Hacking 1979 for some recent developments.) An enumeration of the disciplines that are currently counted as logic, such as standard first order logic, model theory, set theory, and proof theory, does not constitute a satisfactory definition because there are numerous non-standard logics, and because all branches of logic are likely to keep growing. We must know what is constant throughout such

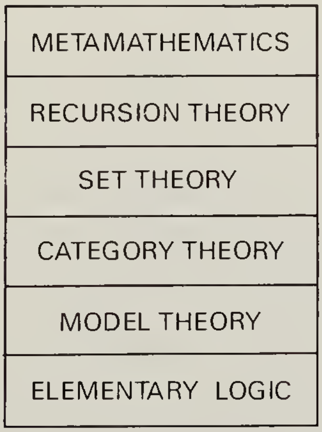


Fig. 1.3. The basement of the mathematical building as of 1985.

changes. The safest answer seems to be this: Logic *stricto sensu* is the basic theory of deduction (a theory completed with model theory), and logic *lato sensu* is the family of all the more general mathematical theories.

(Logic *stricto sensu*, or first order predicate calculus, holds of course a particularly important place in mathematics. Firstly, it has been proved to be both consistent and complete. Secondly, and more important, the predicate calculus is distinguished from all the other mathematical theories in that it presupposes nothing. In fact, once the formation and inference rules have been laid down, everything else follows: elementary logic is its own foundation.)

Being the family of all the most general mathematical theories, logic constitutes both the *foundation* and the *language* of the rest of mathematics. It is the foundation in the sense that every mathematical theory presupposes some branch of logic. (This view has of course a logicist pedigree. Intuitionists regard logic as an offshoot of mathematics, and formalists hold that mathematics presupposes an even more basic theory: that of strings formed out of a finite number of characters.) That this is actually so is shown by axiomatizing mathematical theories in the exhaustive fashion taught us by Tarski (1956). This requires mentioning explicitly the fragment of logic underlying the given theory.

Logic doubles as the basic language of the whole of mathematics, in the sense that every mathematical formula is a particular case of a formula of logic *lato sensu*. (Even Dummett, 1977, who follows Brouwer in holding logic to be derivative, uses the language of logic to write down principles of intuitionistic mathematics.) Thus "$x < y$" is an instance of "Rxy", "$\mathrm{d} \sin x = \cos x \, \mathrm{d}x$" exemplifies "$Fx = Gx$", and "$x$ is a natural number" is a particular case of "$x \in \mathbb{N}$". Caution: The use of elementary logic as a language does not prove that logic is *nothing but* a language. Indeed, whereas languages make no truth claims, theories do make such claims, and logic happens to be a family of theories. Thus the arithmetic falsity "$1 = 2$" instantiates a formula of the predicate logic with identity: the latter is incompetent to assign it a truth value. In sum, logic is the universal language of mathematics (in addition to being the theory of its deductive structure), in the same sense that mathematics is the universal language of science and technology.

The following subsection will illuminate further facets of the question of the nature of logic.

3.2. *Non-Standard Logics*

The logic *stricto sensu* underlying practically all mathematical theories is classical logic, the core of which is the standard predicate calculus with identity. (The two main extensions of this theory, the second order predicate calculus and infinitary logic, occur as well though not as conspicuously.) This is not to say that all mathematicians make explicit use of classical logic. In fact most don't and some, such as Poincaré, Brouwer and the Bourbaki group, have been hostile to logic. Just as most birds probably do not realize that they fly in air, so some mathematicians may not realize that every one of their formulas is an example of a well formed formula of logic, and that every one of their valid reasonings fits into some of the inference patterns canonized by logic.

In addition to classical or ordinary logic there are hundreds of alternative or non-standard systems of logic. (See Gabbay and Guenthner, Eds. 1983–85, Vols. II and III) Among them we recall intuitionistic logic – the only non-standard logic most mathematicians have ever heard about – many-valued logics, modal logics, relevance logics, free logics, quantum logics, and paraconsistent logics. Unlike mathematicians, philosophers of an exact cast of mind are often fascinated by non-standard logics, as can be seen by perusing the *Journal of Philosophical Logic* and the newer *Journal of Non-Classical Logic*. This is not surprising, because most of the non-standard logics have had philosophical motivations, and all of them pose interesting philosophical problems. Suffice it to mention the following: What makes a logical system non-standard?, What are the motivations for the various non-standard logics?, What can one do with them, i.e. what peculiar problems do they help solve?, Are all logics equally worthy, or is there one which is incontestably superior to all others, and What becomes of the unity of science, and even rationality, if there is a plurality of canons?

Before tackling these problems it will be convenient to stress that we shall be concerned mainly with *formal* logics, i.e. theories of deductive inference lacking in ontological, epistemological, or ethical content and presuppositions. This warning is necessary because there are numerous theories which, though called 'logics', are not strictly such because their distinctive basic (primitive) concepts are nonlogical. For example, inductive, epistemic, doxastic, and erotetic "logics" belong in epistemology. In fact, the central concept of the first is that of inductive support, that of the second is certain (as opposed to conjectural) knowledge, that of the third is opinion (or belief), and that of the fourth is problem (or question). Likewise tense logic includes the ontological concept of time, preference logic the psychological and value-theoretic of preference, and deontic logic the ethical notion of duty (or obligation).

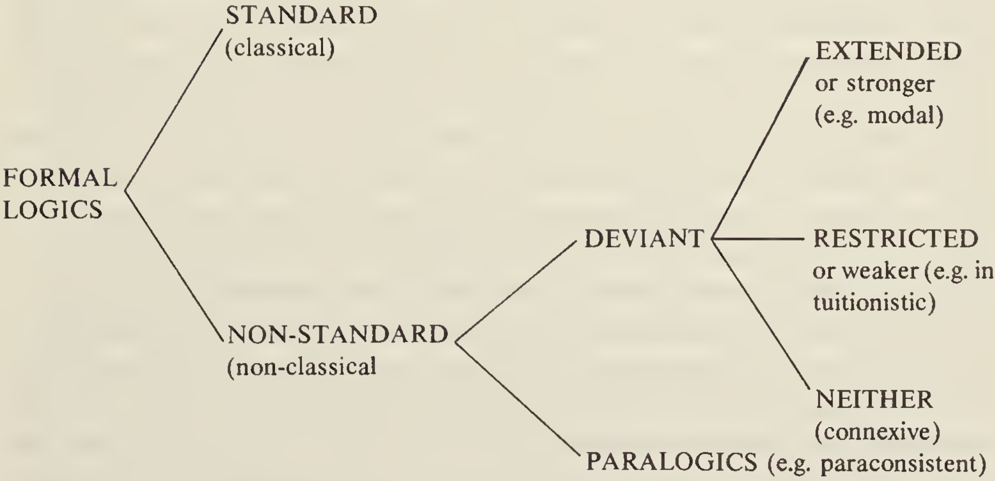
We shall hardly be concerned with these "logics" because they are not purely formal and, far from aiming at elucidating the relation of logical consequence, they all presuppose (classical) formal logic. Besides, we found no use for inductive, epistemic and doxastic logic in Vols. 5 and 6. Nor does one need tense logic in science or in technology, where time – like place – is one among the many "variables" that ordinary logic can handle. (For example, "x happened at t" can be formalized tenselessly as "Hxt & $t <$ present time".) As for the "logics" of preference and duty, as well as those of choice and decision, we shall look into them in Vol. 8. In conclusion, we shall focus our attention on formal logics or, rather, on some of the philosophical problems they raise.

It will prove convenient to start by classing the large, growing, and rather unruly family of formal logics. Our first partition of this family will be the standard/non-standard dichotomy. A *non-standard logic* is any system of formal logic that either fails to contain some classical "principles" – such as those of contradiction, excluded middle, bivalence, or distributivity – or contains additional laws (axioms or theorems). More on this in a moment.

In turn we divide the non-standard logics into those which which do, and those which do not, include the principle of contradiction "$\neg(p \,\&\, \neg p)$". We call *deviant logics* the former and *paralogics* the latter. And we divide the collection of deviant logics into those which extend classical logic, those which restrict it, and those which are neither contained in nor contain it. (See different classifications in Haack 1978 and da Costa 1982a.)

The collection of *extended* deviant logics, i.e. of systems that include classical logic, is exemplified by any classical modal logic, which contains an additional primitive concept, namely that of possibility. The family of *restricted* deviant logics is by far the more numerous: it includes Aristotle's syllogistics, intuitionistic, quantum, relevance, free, and negation-less logics. Classical logic is a supersystem of every one of them, i.e. it is retrieved by enriching the deviant system with one or more principles. The deviant logics that are neither extended nor restricted are a few curiosities such as the *connexive* logics. (See Anderson and Belnap 1975 pp. 434–452.) Finally the *paralogics*, which do not abide by the principle of contradiction, include paraconsistent, universal, nonsense, and fuzzy logics.

In short, we class formal logics as follows:


FORMAL LOGICS
STANDARD (classical)
NON-STANDARD (non-classical
DEVIANT
PARALOGICS (e.g. paraconsistent)
EXTENDED or stronger (e.g. modal)
RESTRICTED or weaker (e.g. intuitionistic)
NEITHER (connexive)

So much for the bewildering variety of non-standard logics. Let us now deal with the first question in our initial list, namely *What is a non-standard logic*? or, put in methodological or pragmatic terms, *How can we recognize logical heterodoxy*? A while ago we gave an apparently simple answer, to wit, a non-standard logic is one that differs from classical logic in at least one of the traditional principles (axioms or rules). The simplicity of this answer is, like all simplicity, deceiving, for there are several alternative formulations of classical logic, and they may not all be strictly equivalent. (I.e. some of them may not systematize exactly the same body of knowledge.) In other words, it is not at all sure that mere inspection of a set of postulates (axioms and rules) of a system of logic suffices to identify it as classical or as non-classical – unless of course one decides on a given classical system as the canonical one. (In this case Tarski's 1965 formalization of predicate logic would be an obvious candidate.)

If form or syntax is not an unambiguous indicator of logical orthodoxy (or heterodoxy), what about interpretation or semantics? Take the case of the "existential" quantifier. As we saw in Sect. 3.1, "$(\exists x)Fx$" can be assigned different interpretations, such as "There are Fs", "Some individuals are Fs", "Some substitution instances of Fx are true", "Individuals with property F can be found", and "An example of an F has actually been constructed". The latter interpretation is favored by the intuitionists, who interpret "p" as "We can confidently assert p", or "p is provable". In fact intuitionistic logic does not differ significantly from classical logic with respect to form: the main difference resides in the interpretation. (However, that small syntactical difference in the underlying logic generates a momentous difference between intuitionistic and classical mathematics, if only because the former does not allow for proofs by *reductio ad absurdum.*) That difference in interpretation or semantics derives from a profound difference in the conception of the nature and role of logic. Whereas to the classical logician logic codifies valid inference, to the intuitionist logic is a means for attaining certain or infallible knowledge. So, in this case the difference is of a philosophical nature. Unfortunately not even a difference in interpretation suffices to identify unorthodoxy, especially when marginal, because – as recalled above – every classical formula can be interpreted in different ways.

How then can we recognize logical heterodoxy? Miró Quesada (1983) has proposed a suitable criterion, which is historical or pragmatic as well as comparative: A logical system is the more heterodox, the more removed it is from mainstream research (i.e. the more it diverges from the evolving tradition). This criterion serves for the synchronic comparison of logical theories but, of course, it does not apply to diachronic comparisons. (Current classical logic might have been regarded as non-standard by Aristotle and perhaps even by Frege.) By this criterion intuitionistic logic, which includes the principle of contradiction, and which underlies some mathematical research, is a mild breeze by comparison with the tornadoes of paralogics. The great majority of mathematicians and philosophers regard the principle of contradiction as *the* over-arching principle of logic, and are therefore inclined to regard paralogics (in the improbable case they have heard of them) not just as deviations from orthodoxy but as aberrations. They are not afraid of novelty but they cannot stand anarchy. We shall come back to this point.

Our second question is that of *motivation*. We all know that the initial motivation for the inception of mathematical logic in the mid-nineteenth century was to update and exactify logic by unearthing, working out, and systematizing the inference patterns actually employed by mathematicians, who had outstripped the ancient and medieval logicians. That motivation was then strictly mathematical; so was the outcome, namely the mathematization and enormous enrichment of classical logic. (Few if any philosophers took notice of this revolution in logic during half a century.) In contrast, the motivations of most non-classical logicians have been philosophical. Let us take a look at a few typical cases.

By far the best known nonstandard logic is the intuitionistic system founded by Heyting. The motivation for it was to formalize and systematize the inference patterns allowed by mathematical intuitionism, one of the main classical philosophies of mathematics. Hence intuitionistic logic inherited the philosophical theses of mathematical intuitionism. Let us recall a few of them. (For a more detailed critical examination see Bunge 1962a.)

Two basic philosophical tenets underlying intuitionistic logic are (*a*) the *pragmatist* maxim "To be is to make something or to be constructed", and (*b*) the *verificationist* (or *operationist*) doctrine of meaning: "The meaning of a statement is the way it is justified (e.g. proved)". These two philosophical views are the foundation of the principle of *constructivity*: "Every mathematical or logical statement expresses the result of a construction" (Heyting 1956a p. 223). Nonconstructive concepts, statements, and proofs, are consequently to be avoided if not outright rejected. Thus where the classical logician might rest satisfied with proving a disjunction, the intuitionist wants to make sure that at least one of the disjuncts be either self-evident or provable. This holds even when one of the disjuncts is the negate of the other: intuitionists neither assert nor deny the excluded middle principle: they just do not use it. Indeed they reject, quite rightly, the idealistic dogma that every proposition *is* true or false regardless of whether anyone has taken the trouble of checking it. And they do not care for existence axioms or theorems unless accompanied by algorithms or by examples. Nor, finally, do they admit indirect proofs – i.e. proofs by *reductio ad absurdum* – which are probably the most numerous (because the easiest) in logic and mathematics.

A quick way of getting a feeling for intuitionistic logic is to look at the manner in which it handles propositions and their denials. To an intuitionist the propositional symbol 'p' should be read 'p is either self-evident or provable', and '$\neg p$' should be interpreted as 'p is either counter-intuitive or undemonstrable'. He admits the law "$p \Rightarrow \neg\neg p$" because he interprets it in an intuitive methodological manner: "If p is demonstrable, then it is not demonstrable that p is not demonstrable". On the other hand he does not admit the converse, namely "$\neg\neg p \Rightarrow p$", because the fact that it is not demonstrable that p is not demonstrable implies nothing about p. None of this affects the classical logician, who feigns that p, $\neg p$, $\neg\neg p$, and the rest, are eternal objects independent of our ability to prove or disprove them. In particular, the classical logician does not feel compelled to construct p in order to assert p.

The restrictions imposed by the principle of constructivity have proved to be double-edged. On the one hand they have prompted the invention of a number of new constructs as well as of direct proofs, which have enriched mathematics. On the other hand this enrichment has not been compensated for by the massacre of classical results cherished by the vast majority of pure and applied mathematicians. Among the uncounted results condemned by intuitionism, at some time or other, as either meaningless, uprovable, or false, are the following: (*a*) every real number is positive, negative, or zero; (*b*) if a and b are real numbers, and $ab = 0$, then either $a = 0$ or $b = 0$; (*c*) every set is either finite or infinite; (*d*) every function continuous in a closed interval has at least one maximum – an existence statement; and (*e*) another existence statement: if a continuous function has a negative and a positive value in an interval, then it has at least one zero in the same interval: see Figure 1.4. Paradoxically, the constructivist constraint has excised from mathematics a great number of intuitive results.

Another philosophical motivation of intuitionistic logic is, of course, the variety of infallibilism known as *intuitionism*: "A mathematical construction ought to be so immediate to the mind and its results so clear that it needs no foundation whatsoever" (Heyting 1956b p. 6). Accordingly one should choose as basic notions the most immediate and obvious ones, such as that of a natural number. A first objection is that intuitability and clarity are subjective or personal, so that they cannot be used as objective criteria for the choice of the basic (primitive) concepts of a theory. If we were to allow our personal intuition to judge the worth of logical or mathematical constructs, then most of mathematics (including intuitionistic mathematics)

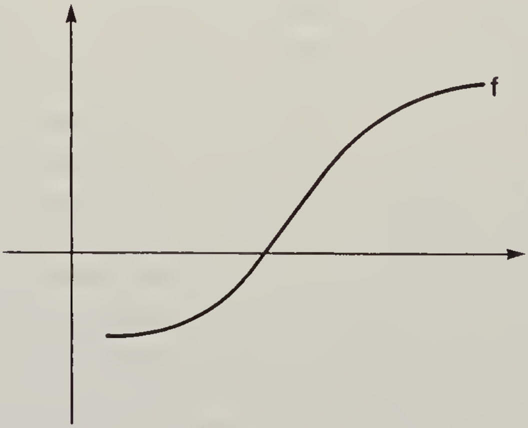


Fig. 1.4. The intermediate value theorem. This existence statement has no place in intuitionistic mathematics because it is not accompanied by a constructive procedure for determining the exact point at which the curve crosses the x-axis. Yet it is surely intuitive and, moreover, of daily use in science and technology.

would be condemned. Remember that mathematics and science start where ordinary experience and intuition leave off (Vol. 6, Ch. 14, Sect. 2). A second objection is that, like it or not, as a matter of fact most mathematical constructs do have some "foundation" or other, i.e. they presuppose others – e.g. because they are built out of more basic constructs. The only foundation-free branch of formal science is ordinary logic, whether classical or intuitionistic.

Finally, it is interesting to note that intuitionism is infallibilist with regard to mathematics but not with regard to logic. In fact according to Brouwer the laws of logic are neither a priori nor eternal: they are corrigible hypotheses rather than immutable regulative principles. The history of logic confirms this thesis up to a point, just as the history of mathematics refutes the intuitionist dogma that there are self-evident and absolutely secure mathematical principles. All mathematical statements and proofs ought to be viewed in principle as capable of being corrected, in particular turned more rigorous. Whether or not the principles of logic should be regarded in the same vein, shall be discussed below.

However, we should draw a distinction that escapes intuitionists, radical empiricists (in particular pragmatists), and vulgar materialists. This is the difference between the *history of the interactions* between logical and mathematical research, on the one hand, and the *logical* relation between logical and mathematical theories considered in themselves, on the other: see Figure 1.5. Logically, mathematics rests on – i.e. it presupposes – logic *lato sensu*. On the other hand the empirical relation between logic and mathematics, i.e. that between the respective research communities, is quite symmetrical. In fact, this relation is one of *mutual and progressive adjustment*.

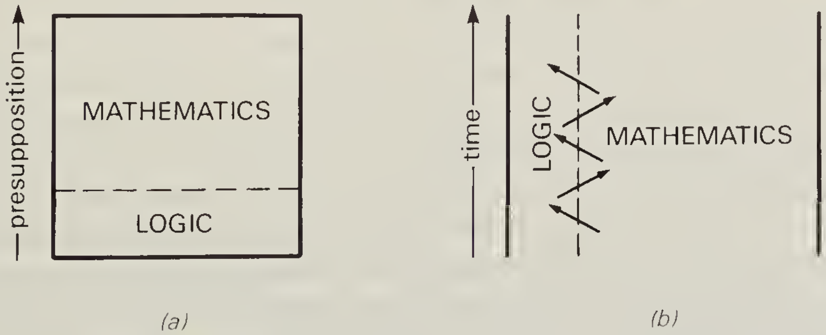


Fig. 1.5. The two relations between logic and mathematics: (*a*) the logical relation of presupposition, and (*b*) the epistemological (and social) relation of mutual inspiration and control.

The validity of mathematics consists in its abiding by the laws of logic, and the validity of logic rests upon its power to analyze and control mathematical research – or rather its deductive aspect. This is not a vicious circle but a virtuous one: it is a process of successive approximations (Bôcher, 1905).

In conclusion, the philosophical motivations of intuitionistic logic are, like those of intuitionistic mathematics, suspect. However, the product is far from worthless: intuitionistic logic is an interesting piece of mathematics. Moreover, properly interpreted – perhaps in epistemological or pragmatic terms, as originally intended – it may yet acquire philosophical significance in addition to showing that it is possible to reason rigorously in non-standard ways.

The next best known nonstandard logic is *many-valued logic* and, in particular, its three-valued version. It has had several motivations (Łukasiewicz 1970). One of them is semantical (or epistemological), namely the view that some propositions are neither true nor false, but just "possible" (or possibly but not actually true or false). In turn, this semantical thesis has a metaphysical substratum, namely indeterminism and, in particular, the doctrine of future contingents, which goes back to Aristotle. Consider the classical example: whether or not there will be a sea battle tomorrow is not fully determined by present conditions, and this quite apart from our inadequate knowledge of the latter. Consequently, the proposition "There will be a sea battle tomorrow" is just as "possible" as its contradictory, namely "There will not be a sea battle tomorrow". So, neither of them should be assigned the value "true". Three valued logic assigns them the truth value $\frac{1}{2}$, or else "indeterminate"; moreover, it introduces important changes in the axioms. (See Rescher 1969.)

In Vol. 2, Ch. 8, we chose to keep classical logic and give up instead the assumptions that every proposition has a truth value (whether we know it or not), and that the only possible truth values are 0 and 1. This strategy allows us (*a*) to abstain from assigning truth values to the above propositions about the future sea battle (or any other indeterminate events), (*b*) to handle half-truths, such as "This book is thoroughly wrong", by introducing the notion of graded or partial truth, and (*c*) to retain ordinary logic as the basis of mathematics, as well as the common core of all scientific theories, formal and factual. The latter is no small advantage compared to the staggering cost of rebuilding the whole of mathematics and science to conform to some system of many-valued logic. (The same holds, of course, for any other deviant logic.) But, again, this does not detract from the intrinsic interest of many-valued logic.

Our third example of a nonstandard logic is *modal logic* (Lewis and Langford 1959, Hughes and Cresswell 1968). C.I. Lewis had two motivations for initiating it. One was the wish to form a concept of logical consequence untainted by the so-called paradoxes of implication; the other was to elucidate the notions of necessity and possibility. The logical motivation can be defused by noting that mathematics need not recapture all of our premathematical intuitions. (Besides, it soon became apparent that modal logics themselves are ridden with paradoxes; e.g. the denial of a necessary proposition strongly implies any proposition.) As for the notion of logical consequence, it was elucidated several years later by model theory, which taught us that (*a*) entailment is stronger than implication (i.e., if p entails q, then p implies q, but not conversely), and (*b*) there are two different though related concepts of entailment: the syntactical ($\vdash$) and the semantical ($\models$) relations, both of which exactify the vague notion of logical (or necessary) consequence. (Necessarily, if p then $q =_{df} p \vdash q$. Necessarily, if p is true, then q is true $=_{df} p \models q$.)

What about the ontological motivation for modal logic, i.e. seeking to elucidate the notions of physical (or ontic) possibility and necessity? Unfortunately modal logic does not do this job either. As we saw in Vol. 3, Ch. 4, Sect. 2, in factual science a fact is regarded as being possible only if it fits some objective (e.g. natural) law – and such law need not be causal. Therefore a science-oriented ontology has no more use for modal logic than mathematics. Still, the failure of modal logic to elucidate the notions of entailment and of real possibility does not detract from its interest as a rich family of rigorous and ingenious mathematical systems. Moreover we cannot rule out the possibility that, properly interpreted, modal logic may still acquire some philosophical significance.

Next in our list comes *relevance logic* (Anderson and Belnap 1975). The motivation for this system is to avoid the admixture of heterogeneous universes of discourse permitted by classical logic (though prohibited by totle). We saw one example of this only one minute ago: a falsity implies anything and, in particular, *ex contradictoriis quodlibet*. Another, and far more frequent, source of irrelevance is the principle of addition: $\ulcorner p \vdash p \vee q \urcorner$, where q need not be related to p. Relevance logic attempts to reform classical logic so as to respect relevance, i.e. to stick to the original universe of discourse. However, it does not succeed, because it keeps a major offender, namely the principle of addition. And it does well to keep this principle, for we need it in open (e.g. developing) contexts, where we wish to retain the freedom to relate subject matters that, though at first blush mutually foreign, may prove to be intimately connected. After all, relevance is contextual not absolute.

I submit that the very project of building a relevance logic is misguided because the concept of relevance is not only contextual but also semantical rather than syntactical. Relevance is a matter of reference, and relevance logic does not contain a theory of reference (for which see Vol. 1, Ch. 2). Relevance can be controlled by alternative means. For one thing, if we need to secure relevance (or homogeneity of the reference class), we can do so by axiomatizing the body of knowledge concerned. Indeed, axiomatization specifies in advance the universe of discourse and it guarantees reference conservation (Vol. 1, Ch. 2, Sect. 4.2.). Secondly, if (in an open context) we suddenly encounter a total stranger to our originally intended referents, then we may suspect either of two things: that the irrelevance is real, or that it is apparent. In the first case irrelevance will be a product of some contradiction that has gone unnoticed. (By the way, this shows the usefulness of the "paradox" *ex contradictoriis quodlibet.*) But the irrelevance may be only apparent: we may have discovered a previously unsuspected link between two reference classes and their corresponding research fields. Classical logic may help us in the first case, relevance logic in neither. (See further objections in Copeland 1980.)

Let us now peek at the philosophical motivations for the various "logics" of change. One of them is *temporal* (or *tense*) "logic" (Prior 1967, Rescher and Urquhart 1971), disqualified earlier as a logic proper because it contains the ontological concept of time. One such system of temporal logic is nothing but the classical predicate calculus enriched with an extralogical predicate interpreted as "is an instant of time". Another such system may be construed as the theory of partially ordered sets (which presupposes classical logic), in which the basic set is interpreted as the collection of (absolute) instants, whereas the order relation is interpreted as "(absolutely) prior to or simultaneous with". None of these theories conceives of time as relative to some reference frame or even as a continuous parameter; hence temporal "logic" is of no interest to scientific ontology. In sum, temporal "logic" is a miniontology of no interest to either philosophy or the foundations of science: it is but an academic exercise in exact philosophy.

(As for the linguistic aspect of the question, it should be noted that every tensed sentence can be transformed into a tenseless one by an appropriate use of the concept of time – paradoxical though this may seem. Thus "*a* came before *b*" can be converted into "*a* precedes *b* in time", or "*a* is prior to *b*" – perhaps with the addition of the clause "relative to frame *c*". These tricks do not expel becoming from language: they just deprive the present, which is egocentric, from its privileged position. Hence they allow us to treat all instants of time, relative to a given reference frame, on the same footing. This results in greater objectivity and universality – which is why it is the method adopted in science and technology.)

However, the most radical of the "logics" of change is not a miniontology, like temporal "logic", but *paraconsistent logic* (da Costa 1980, Arruda 1980, Routley 1980). Its peculiarity is that the principle of contradiction is not a logically valid schema in it, whence it would be the "logic" underlying inconsistent theories, such as naive set theory. On the other hand most other logical principles, including the excluded middle and the modus ponens, are valid in paraconsistent logic – which is itself consistent.

One motivation of paraconsistent logic is the wish to avoid the classical law *ex contradictoriis quodlibet* (a contradiction entails anything). This permissiveness is embarrassing because (*a*) it condones *trivial* systems, i.e. systems all the formulas of which are theorems for being entailed by a single contradiction; and (*b*) it makes it hard to *evaluate* axiom systems by checking only their consequences. However, there are two ways of avoiding that law without giving up rationality. One is to prove the consistency of the axiom system of interest or, even better, to compress all of its axioms into one. (This is one of the reasons for preferring physical theories based on a single variational principle.) Another is to adopt intuitionistic logic, for it demands that the antecedent of every conditional be provable. Either procedure, though expensive, keeps us in business, whereas adopting paraconsistent logic amounts to declaring bankruptcy. Besides, as we saw a moment ago, the law *ex contradictoriis quodlibet* is useful in open contexts: it helps us suspect irrelevance from contradiction and conversely. In sum, there is no logical justification for adopting paraconsistent logic.

Another motivation for paraconsistent logic is ontological, namely the conflation of logic with ontology and the wish to accomodate dialectics, according to which change is contradictory, so that some contradictions would be true of our changing real world (da Costa 1980, 1982b, Routley 1980). In particular, ultralogic – the most ambitious of all paraconsistent logics – would be about everything thinkable, whether conceptual or material, possible like a black hole or impossible like a round circle: "It provides a canon for reasoning in every situation, including illogical, inconsistent and paradoxical ones" (Routley 1980 p. 893). Thus the classical rationalistic illusion that reason is self-sufficient ends up by destroying rationality.

We have argued before (Sects. 1 and 2) that logic and pure mathematics are only about constructs, not material things. Consequently a universal logic or ontology *à la* Meinong, dealing with anything thinkable, must be either utterly trivial or false: constructs and concrete things have no laws in common. (See also Vol. 3, Introduction, Sect. 2.) We have also argued (Vols. 5 and 6) that the study of the real world involves empirical procedures: that no purely a priori theory can be hoped to be true of concrete things. Hence in our perspective any attempt to build an a priori theory encompassing objects as heterogeneous as inferences and quantum jumps is wrongheaded.

As for dialectics, we have argued elsewhere (Bunge 1975a) that it is not an adequate theory of change and novelty – except insofar as it stresses the centrality of both – if only because it regards change as contradictory. (This view seems to have originated in the Eleatic analysis of motion. The arrow was said *to be and not to be* at a given place at a given time. Science analyzes motion, and change in general, in a totally different fashion, namely by *blending* the concepts of thing (or being) and change (or becoming) rather than opposing them. Thus we say that the arrow *moves* with velocity v when passing through place x at time t. We assume, more generally, that every thing is at a given time, relative to some reference frame, in some state (which may be composite). And in turn a state may be construed as a point or stage in some process or other: Vol. 3, Ch. 5). True, some processes are mainly conflictive or dialectical; but others are mainly synergic or cooperative, and still others are neither or are a combination of several basic modes of change. In any event change (*a*) has nothing to do with logical contradiction, and it is not within the domain of the logician, and (*b*) it is representable by factual theories which have classical (at most intuitionistic) logic built into them. In short, there is no legitimate ontological or scientific rationale for doing paraconsistent logic. On the other hand there is a powerful reason for avoiding it, namely the defense of rationality.

Still, the nagging question remains: *Why* should rationality involve the principle of non-contradiction? Certainly not because contradictions are unthinkable: we often incur them. Nor because they are unintelligible: we all understand a contradiction such as "The world is small and big". Nor because they are meaningless: in fact contradictions are just as meaningful as tautologies. (Our result on the meaning of a tautology, in Vol. 2, Ch. 7, Sect. 2.1, holds also for its negation.) Nor, finally, because contradictions are false: it would be up to us to decree them true. I submit that we value non-contradictoriness for extralogical reasons. Firstly because we cherish truth, which in turn involves consistency: think of the prospects of finding out the truth about the author of a crime on the sole strength of the accounts of two witnesses, one of whom swears that *A* did it, and the other that *A* did not do it. Secondly because, when interpreted in practical terms, contradiction leads to inaction: indeed it is impossible to do *A* while at the same time not doing *A*. In conclusion, we value consistency mainly for practical reasons. By the same token dialectics, far from being "the algebra of revolution" (Lenin), condemns us to inaction and confusion. So much for paraconsistent "logic" and its motivations.

Another system that we have characterized as paralogical is *fuzzy logic* (Zadeh 1975). Its goal is to describe ordinary reasoning, which is for the most part imprecise because it contains fuzzy predicates, such as "long", and plausible (e.g. analogical) inferences rather than valid ones. Indeed, far from being normative like classical logic and all of the deviant logics proper, fuzzy logic "aims [...] at an accomodation with the pervasive imprecision of human thinking and cognition" (Bellman and Zadeh 1977 p. 106). This is of course a fascinating problem for cognitive psychology and epistemology (Vol. 5, Chs. 5 and 6). However, it is misleading to call that study 'logic' if the whole point of it is to *legitimate* fuzzy notions and invalid inferences instead of discovering and correcting them. (On the other hand fuzzy "logic" is a priori in so far as it makes no use of empirical procedures for finding out how people actually reason.)

Fuzzy logic is defeatist: it "represents a retreat from what may well be an unrealizable objective, namely the construction of a rigorous mathematical foundation for human reasoning and rational behavior" (Zadeh 1975 p. 426). The historical evidence points to the opposite course: it points to the advancement of rigor – though not without some temporary setbacks. And even if it did not, it is the duty of the logician to reduce fuzziness and invalidity rather than to portray them or shove them under an attractive formal rug. What is the point of formalization unless it is to substitute precision for fuzziness, and validity for sloppy inference? What is the function of the torch but to help us find our way in the dark?

Finally, let us examine the motivations for *quantum logic*. But first we must dispel a common confusion, that between quantum *logics* and the peculiar *algebras* used in quantum theory. This confusion stems from Birkhoff and von Neumann (1936), who conflated the concept of proposition with that of a projection operator, only because both have two possible values, namely 0 and 1. This confusion is facilitated by adopting an "operational" redefinition of a proposition that is likely to elicit a smile from any logician. One such redefinition, originally suggested by the physicist Jauch, is this: "Propositions are observables the measurement of which can only give one of two values conventionally called *yes* and *no*. Hence, a statement is a proposition only if an instrument, by means of which its content can be verified, can be conceived. Conventional quantum mechanics associates projectors to propositions" (d'Espagnat 1973 p. 717, also 1975). We shall steer clear from such confusion and shall deal exclusively with quantum logics proper, i.e. the systems of nonstandard logic that have sometimes been claimed to underly quantum theory.

Two main motivations have been given for devising quantum logics. One is the fact that quantum theory contains mutually incompatible (or "complementary") propositions, and therefore propositions that cannot be truly asserted jointly. The other is the claim that certain experiments, notably those involving the diffraction of electrons and the like, refute the distributive law for propositions. (See e.g. Friedman and Putnam 1978.) To understand the first motivation let us recall that quantum mechanics represents dynamical variables by operators, and that some of these do not commute with one another: i.e. if $\hat{A}$ and $\hat{B}$ are such operators, it may be the case that $\hat{A}\hat{B} \neq \hat{B}\hat{A}$. In this case the proposition "The value of A is a" is incompatible with (or complementary to) "The value of B is b". (For example, "The value of the x-component of the angular momentum is $m\hbar$" and "The value of the y-component of the angular momentum is $m\hbar$" are not both true of a given thing in a given state at a given moment.) The notion of complementarity may be regarded as a generalization of that of contradiction – provided one does not forget that the former is contingent whereas the latter is logical. Indeed that two propositions are mutually complementary is not a matter of form but of matter: it represents a feature of reality. If A and B fail to *have* simultaneous sharp values, then we cannot truly attribute them such values. On the other hand we can attribute them simultaneous distributions of values. And the more peaked the distribution of A, the less so that of B, and conversely.

Now, it would make no sense to build a system of logic that excludes mutually complementary propositions, as quantum logic is sometimes said to do. Logic must be comprehensive enough to allow us to state mutually complementary (and even contradictory) propositions. It is up to empirical research to find out which if any of the components of such complementary propositions is true. This holds not only for quantum theory but for every research field and every application of logic. It happens even with reference to daily life. Thus the propositions "She is swimming now" and "She is typing now" can never be both true for the same *she* and the same *now*, but this is a material constraint that does not affect logic. All we have to do when encountering mutually complementary propositions is to abstain from stating them jointly. This, rather than stating that the underlying logic is non-Boolean, is the reasonable attitude to take. To do otherwise is to confuse logic with ontology or even physics.

As for the claim that quantum mechanics violates the distributive law for propositions, let us illustrate it with the help of an example used above. Consider the propositions

$A = \ulcorner$The x-component of the angular momentum of the electron is $m\hbar\urcorner$
$B = \ulcorner$The y-component of the angular momentum of the electron is $m\hbar\urcorner$
$C = \neg B$,

where the electron referred to in all three propositions is a given electron in a given state at a given moment. Now, the distributive law, for propositions is the equivalence

$$A \,\&\, (B \vee C) \Leftrightarrow (A \,\&\, B) \vee (A \,\&\, C).$$

Since $B \vee C$ is a tautology in ordinary logic, the *LHS* of the equivalence amounts to A, so it is true if A is. Consider now the *RHS* of the distributive law. In our example $A \,\&\, B$ is false because A and B are mutually complementary propositions. (The corresponding operators do not commute and therefore have no simultaneous sharp values. This, in our realistic interpretation, is why we cannot measure such sharp values on one and the same object at the same time: see Ch. 2, Sect. 4.1.) On the other hand $A \,\&\, C$ is true if A is, for when the x-component of the angular momentum has a sharp value, the y-component does not. Therefore the *RHS* is true if A is. But, unlike the *LHS*, the *RHS* does not simplify to A. Hence it would seem that, indeed, the distributive law fails in quantum mechanics.

Worse, we need not resort to quantum physics to find apparent cases of non-distributivity. Take the case

$A = \ulcorner$She is swimming now$\urcorner$
$B = \ulcorner$She is typing now$\urcorner$
$C = \neg B$.

Here too the *LHS* of the distributivity law simplifies to A, whereas the *RHS* would seem not to. So, the quantum has nothing to do with the apparent failure of distributivity. The latter has to do with an incomplete analysis of the argument.

Indeed, in the above two arguments we have made tacit use of the premise that A and B are mutually "complementary" (incompatible), i.e. that

$\neg(A \& B)$. We employed this premise in evaluating the first disjunct of the *RHS* of the distributive law. If this extra premise is taken into account, it turns out that distributivity holds after all. The proof runs as follows

LHS	*RHS*
1. $A \& (B \vee \neg C)$ assumption	1′. $(A \& B) \vee (A \& \neg B)$ assumption
2. A 1 and simplification	2.′ $\neg(A \& B)$ tacit assumption
	3′. $A \& \neg B$ 1′ and 2′.
3. $\neg(A \& B)$ tacit assumption	
4. $\neg A \& \neg B$ 3 and de Morgan	
5. $\neg B$ 2 and 4	
6. $A \& \neg B$ 2 and 5.	

We have proved that the *LHS* and the *RHS* have the same implicates. To prove that these in turn imply the original propositions we proceed as follows.

LHS	*RHS*
7. $B \vee \neg B$ 5 and addition	4′. $(A \& B) \vee (A \& \neg B)$ 3′ and addition.
8. $A \& (B \vee \neg B)$ 2 and 7.	

In short, the distributive law is seen to hold when the tacit assumption concerning the mutually "complementary" propositions is rendered explicit. Hence the second motivation for playing with logics characterized by non-distributivity, in particular quantum logics, evaporates as well.

The upshot of the preceding discussion is that quantum logics are not well motivated. (Moreover they are mathematically suspect as well, for every lattice endowed with implication, i.e. every Brouwer algebra, is distributive: see Rutherford 1965 p. 74.) We could have arrived at our result in a far quicker way, e.g. by pointing out that (*a*) so far nobody has ever employed a quantum logic to derive a correct result in quantum theory, and (*b*) work in quantum logic over half a century has produced no new scientific results. The explanation for this lies in the fact that quantum theories contain only mathematical theories the underlying logic of which is classical, as can be seen only by axiomatizing quantum mechanics (Bunge 1967c). So much, for the time being, for the motivations of non-standard logics.

Let us now tackle the third problem in our agenda, *viz.*, *What can be done with deviant logics that cannot be done with standard logic*: i.e., which kinds of new problems can they help solve? At first sight the answer is simple enough: A logical system that is restricted (extended) with respect to standard logic allows one to pose a subset (superset) of the problems that standard logic can handle. This answer is correct as far as the *logical* problematics goes. For example, a negation-less logic cannot handle any problems involving negation.

However, that simple answer is wrong with respect to the *mathematical* problematics. Indeed, a restricted logical system, provided it is not too poor, allows for more freedom to mathematical inventiveness because it imposes less restrictions. For example, intuitionistic logic, by not including the theorems of double negation, excluded middle, and de Morgan, underlies certain mathematical theories disallowed by ordinary logic. (An important recent example is topos theory.) However, such enrichments in some departments of mathematics are paid for by severe curtailments in others. Therefore it would be rash to either impose or outcast intuitionistic logic: a moderate pluralistic foundational policy seems advisable. See Sect. 5.1.

The situation with regard to the systems of extended logic is the dual of the preceding: they allow one to pose more logical problems but, precisely because they are richer (more specific), they restrict mathematical inventiveness. Thus modal logic is more comprehensive than classical logic, but mathematicians have found no use for it except as an object of metamathematical study. And even the worth of the additional logical riches gained by the extended logical systems is doubtful. Recall that modal logic and other extended logical systems failed to solve the problems they set out to solve. Those problems are best tackled elsewhere – e.g. in model theory, semantics, epistemology, or ontology.

In short, the only deviant logic that has proved to be of use to mathematics is intuitionistic logic – as long as it is separated from its philosophical concomitants. The remaining non-standard logics have proved useless and have no sound philosophical motivations. This rather negative evaluation of deviant logics does not entail that they have been barren. On the contrary, they have sparked off some interesting work in algebra (e.g. Rasiowa 1974). And, more important to us, they have helped us – albeit at a staggering price – to better understand the nature of logic and even of rationality. In particular deviant logics have taught us that (*a*) whatever else it may be, formal logic is not the science of the laws of thought; (*b*) the logical principles, far from being eternal truths, are our own brain children and, as such, subject to some change in the course of history – though we must keep some invariants if we are to continue to study the same subject and retain rationality; (*c*) it is possible to reason in numerous non-classical ways – e.g. intuitionistically or modally – without incurring contradiction; (*d*) however queer a logical theory may be – even if it casts off the principle of contradiction – it can be formalized, whence formalization does not guarantee exactness, consistency, or usefulness: it only helps attain them; (*e*) formalization is necessary but insufficient to handle topic-dependent problems, such as those of reference (which motivated relevance logic) and change (which motivated temporal logic).

In view that all of the logics proper are equally rigorous, i.e. precise and consistent, it might be thought that our preference for the two main systems – classical and intuitionistic logic, and particularly the former – is sheer conservatism. Granted: if rigor (formal correctness) were the only choice criterion, then we ought to be logical relativists or pluralists, as Rescher (1977 a Chs. xiii and xiv) has argued with his usual force. But in fact mathematicians, the masters of contemporary logic, are more exacting: they demand that a system of logic *stricto sensu* satisfy all of the following admissibility conditions:

(i) A system of logic should be precise: it should contain only exact concepts;

(ii) a system of logic should include all the basic operations of formal reasoning (such as negation and conjunction, "existential" quantification and universal quantification), it should contain the modus ponens and – last but not least - the principle of contradiction;

(iii) a system of logic should be able to analyze and systematize the inference patterns that have been found successful in mathematics and in factual science; and preference should be accorded the system which best captures the subtleties of mathematical and scientific reasoning, and which keeps abreast of mathematical and scientific advances;

(iv) no system of logic is admissible which condones fallacies (such as affirming the consequent) or generates insoluble paradoxes (to be distinguished from merely counterintuitive propositions);

(v) no system of logic is admissible which impoverishes mathematics or science to the point of precluding the formulation of well worn theorems or law statements, or which misguides people by suggesting that logic can solve by itself problems in semantics, epistemology, ontology, ethics, or factual science.

I submit that classical and intuitionistic logic are the only known logical theories satisfying all of the above conditions. (Moreover classical logic is the only one that harmonizes with the bulk of mathematics and factual science. This is no mean feat for, if a different logic were to be adopted, the whole of factual science would have to be renovated in order to jibe with this modified mathematics.) Finally, classical and intuitionistic logic are far from being stagnant bodies of knowledge, so the charge of conservatism is groundless. In fact they have been moving so fast these past decades that anyone who got his training in logic twenty years ago, and has not kept up to date, would be unable to read the current issues of the logic journals. In conclusion, unrestricted logical relativism (or pluralism) does not pay: logical monism does, and dualism may.

We are now ready for our final question: *What becomes of rationality if there is a plurality of formally correct systems of logic*? The answer to this question depends critically upon the definition of "rationality" that one may care to adopt. Now, the very first problem we meet when attempting to define the notion of rationality is that it is not unique. Indeed one must distinguish at least the following seven concepts of rationality (Bunge 1985):

(i) *conceptual*: minimizing fuzziness (vagueness, imprecision);

(ii) *logical*: striving for consistency and giving reasons for every claim;

(iii) *methodological*: questioning and demanding proof or evidence;

(iv) *epistemological*: avoiding conjectures incompatible with the bulk of background knowledge;

(v) *ontological*: adopting a self-consistent naturalistic world-view;

(vi) *valuational* (Weber's *Wertrationalität*): striving for goals that, in addition to being attainable, are worth being attained;

(vii) *practical* (Weber's *Zweckrationalität*): adopting means likely to help attain the goals in view.

I submit that these various kinds of rationality are ordered in the way they appear above. (See Mosterín 1978 for the thesis that practical rationality presupposes theoretical rationality.) Accordingly, conceptual and logical rationality are necessary for rationality of any other kind: they compose *minimal rationality*. If we stick by the latter we shall avoid or refine fuzzy predicates, which do not satisfy classical logic; and we shall avoid or remove inconsistency, which hinders the search for truth.

Our definitions of conceptual and logical rationality condemn paralogics – in particular paraconsistent, nonsense, and fuzzy logics – as irrational. Indeed, by condoning fuzziness or inconsistency they break the paramount rule of the logical game. (If you wish your game to be called 'soccer', abide by the rules of soccer.) In particular, paraconsistent logic tolerates empty concepts, such as the set of all the objects that possess and fail to possess a given property. And it tolerates, nay encourages, inconsistent theories, which contain mathematically nonexistent objects (recall Sect. 2.1). (Moreover an inconsistent theory is really a set of at least two mutually incompatible systems, every one of which can be supported by some body of "soft" data.)

In our view paralogics promise only intellectual anarchy: the destruction of all kinds of rationality. The paralogician, dazzled by formalization – which is normally but a means – has lost sight of the job of logic: he has abdicated the logician's duty, which is to prevent reason from going astray, and to help us reduce fuzziness and avoid inconsistency.

On the other hand our definitions of conceptual and logical rationality do not exclude any deviant (but consistent) logics, whether restricted, extended, or neither. All of the deviant logics spare rationality although they draw differently the borders of reason. The problem is to decide whether we need or want a shift in the frontiers at this time. And this is not a logical problem but one involving mathematics, factual science, and philosophy. We now have the means to solve that problem. Our answer is this:

(i) *we do not need* the extended and connexive logics because in some cases they do not solve any significant problems other than those solved by standard logic, in others they do not solve correctly the problems they were intended to solve, and in still other cases they attempt unsuccessfully to grapple with problems belonging to other departments, notably semantics;

(ii) *we do not want* the restricted logics – except possibly the least restrictive of all, namely intuitionistic logic – because they mutilate important branches of the tree of knowledge;

(iii) *we need and want* to stick to classical logic, or at least to intuitionistic logic, because these systems

(*a*) satisfy all of the admissibility requirements listed a short while ago - essentially the ability to exactify, analyze and systematize all the successful inference patterns;

(*b*) have proved useful not only as tools of analysis and canons of discussion, but also as tools for reconstructing intuitive mathematics in a rigorous way, and occasionally also for building new mathematical theories (as was recently the case with non-standard analysis, an offshoot of model theory, and with topos theory, an offshoot of category theory);

(*c*) satisfy the requirements of conceptual, logical, epistemological and methodological rationality.

At the time of writing, standard (classical) logic has the additional advantage that it underlies the bulk of mathematics and therefore of all the factual sciences. (Intuitionistic logic underlies only a comparatively small portion of mathematics, and it has yet to be adopted by any factual theories.) And logical unity is of paramount importance because it makes it possible for the various research fields to interact with one another – an interaction that promotes the unity (or systemicity) of human knowledge. The moment one branch of learning were to adopt a non-standard logic it would cease to import and export ideas: it would become insular and so, eventually it would degenerate into a pseudoscience. (Recall condition (xi), in Sect. 1.1, for a branch of knowledge to be scientific.) For example, if quantum mechanics were to adopt a logic of its own it would cease to interact with classical physics and would thus become untestable. (This should suffice to write off quantum logics.) For the time being intuitionistic logic does not comply with the unity of knowledge requirement: it is marginal. This situation is tolerable only as a transient state. Eventually the whole of mathematics will have to be built on the basis of either classical or intuitionistic logic – or of a third still unknown system.

To conclude. Non-standard logic is a variegated field of exact if largely idle research. Further developments in it are unavoidable given the curiosity and lust for generality of mathematicians, as well as the fact that many philosophers with an exact cast of mind have a tendency to tackle epistemological, ontological and ethical problems in an a priori manner, producing "logics" of change, belief, preference, and much else. However, after nearly a century, non-standard logic can exhibit but one useful system, namely intuitionistic logic. The remaining non-standard calculi are just intellectual games that have failed to solve any philosophical, mathematical or scientific problems.

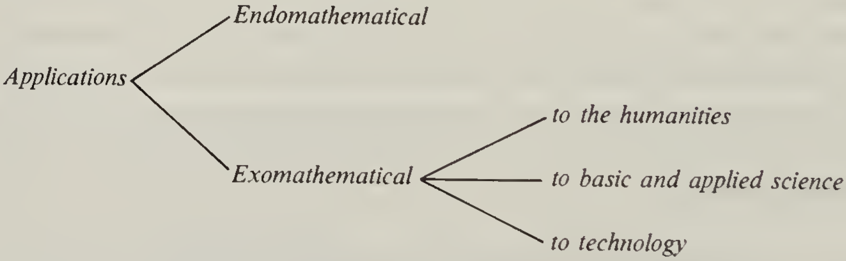
4. PURE AND APPLIED MATHEMATICS

4.1. *Applications of mathematics*

Every mathematical idea can be worked out either for its own sake or in view of some application. Moreover, in principle every mathematical idea can be used in contexts other than that in which it originated. Indeed, mathematics and philosophy are the most diffusive or portable of all fields of knowledge.

However, not all applications of mathematics are to extramathematical fields. Thus number theory, which originated in counting, developed into a branch of pure mathematics which was eventually applied to geometry and other branches of pure mathematics. Group theory, which originated in algebra, was later applied to geometry and factual science, mainly physics. On the other hand geometry, which originated in land surveying, became a branch of pure mathematics which in turn found uncounted applications in factual science and technology. Likewise the infinitesimal calculus, which was partly conceived as the mathematics of motion, evolved into a theory in its own right and was eventually applied not only in natural science and technology but also in economics. There is a constant give and take between pure and applied mathematics.

The above examples suggest classing the applications of mathematics into *endomathematical* and *exomathematical.* Thus the use of set theory to reformulate mathematical analysis in a more rigorous (and less intuitive) fashion is an example of an endomathematical application. On the other hand the following are instances of applications of mathematics to other disciplines: (*a*) to the humanities: using the notion of a function to elucidate that of reference, and that of a semigroup to exactify that of exact language; (*b*) to factual science: using manifold geometry in the theory of gravitation, and probability theory in the theory of social mobility; (*c*) to technology: using queueing theory to design telephone exchanges, and probability theory to design quality control procedures. In sum, we have the following systematics of the applications of mathematics.



Because some applications of mathematics are done within mathematics – as is the case with the algebraic treatment of logic – the collection of applications of mathematics is not identical with what is usually called ‘applied mathematics’. (I.e. some applications of mathematics remain within pure mathematics.) Unfortunately the expression ‘applied mathematics’ is ambiguous. In some contexts it stands for the study of mathematical problems that happen to arise in research fields other than mathematics, though divested of their factual content. This is the case with queueing theory, the general theory of control, and potential theory. (The latter, a part of mathematical physics, was born from the classical theory of gravitation, which is in turn a branch of theoretical physics.) In other cases 'applied mathematics' designates the collection of exomathematical applications of mathematics. In this sense exact philosophy and a mathematical model of an ecosystem may be said to belong to applied mathematics.

The difference between the two senses of 'applied mathematics' is one of reference and of attitude or interest. In the first case the applied mathematician is interested in mathematics *per se*, and consequently he produces mathematical theories or theorems that have no precise reference. (For example automata theory, though initially motivated by computer science, can also be worked on as a branch of pure mathematics.) In the second case the applied mathematician is not primarily interested in mathematics but uses the latter as a tool for thinking about some nonmathematical object – such as the concept of truth, or a thermonuclear reaction, or a change in the frequency of a genetic trait – or for controlling some natural or social process. He is a humanist, scientist, or technologist, who handles mathematics as a tool. Let us concentrate on this second sense of 'applied mathematics'.

Consider first the most basic branch of mathematics, namely logic *lato sensu*. Applied logic is of course the application of the concepts and principles of logic to nonlogical problems. There have been two extreme stands with regard to the scope of such application: nihilism and panlogicism. The nihilists, echoing Pascal's distinction between the *esprit de géométrie* and the *esprit de finesse*, believe that logic is just shorthand (symbolization), or trivial routine like doing sums (Ryle), or a caricature of ordinary language. Ordinary language philosophers and irrationalists have made these claims.

All three claims are grossly in error. The first, because logic is a set of theories not just a symbolism. The second, because new, ever more difficult problems, keep cropping up in logic. The third, because logical analysis has been able to exactify uncounted coarse ordinary knowledge notions and reasonings. Only logic (and more generally mathematics) is imbued of the *esprit de finesse*. Ordinary knowledge and its medium, ordinary language, embody the *esprit de vulgarité*.

The optimists concerning the scope of logic believe, on the other hand, that the logician *qua* such is competent to approach any problems, conceptual or empirical, philosophical or scientific. There are still logicians daring enough to tilt at any problems with the sole help of the logical lance: witness the unchecked proliferation of deviant extended logics dealing with such factual items as becoming. (See Sect. 3.2.) Still, some philosophers realize that such attempts are doomed to failure for lack of contact with factual science.

I submit that the main exomathematical applications of logic are to philosophical problems, and even so to problems of a restricted kind, namely of philosophical analysis – in particular the analysis (or exactification) of philosophical concepts and propositions, and the analysis of reasoning, dialogue, and disputation. On the other hand philosophical synthesis, or the construction (or reconstruction) of philosophical theories in ontology, epistemology, ethics, or other fields, should be out of bounds because it requires not only more powerful formal tools (such as abstract algebra and topology), but also some substantive knowledge (e.g. in physics or in biology). Thus a theory of mind that were to use only the predicate calculus and our commonsense knowledge about mental phenomena would be hopelessly shallow and out of touch with the science of mind. Far from being a contribution to knowledge, it would merely be an intellectual pastime.

We adopt then a moderate stand with regard to the scope of applied logic. But even within this moderate stand one may feel that there are limits, or that there are none, to the exactification of concepts. Pessimists feel that there are certain inherently confused or vague notions bound to resists all efforts at mathematization. One such notion would be that of impetus, the cornerstone of a non-Aristotelian dynamical theory that flourished between the 6th and the 16th centuries (Koyré 1943). But this thesis is debatable. After all, the impetus – which was supposed to keep bodies moving after they had been set in motion – resembles our modern concept of inertia (mass). So, it is conceivable that a competent mathematician could have exactified it – though not in isolation but as a member of a hypothetico-deductive system.

Any interesting notion is worth an exactification effort. (But of course it won't pay to use heavy artillery to kill flies – e.g. to use differential equations to account for yearly economic statistical figures.) To drop an interesting concept for being vague is defeatism, because all new concepts are born fuzzy. And to wait for it to be clarified without the help of mathematics is no better than believing in miracles. The right policy with regard to the bounds of exactification is summarized in

RULE 1.1. Try to exactify every interesting concept, if necessary embedding it in a mathematical theory. And push the exactification effort as long as it is worthwhile.

Not everyone believes in exactification. In fact its opposite, or fuzzification, is now quite trendy among applied mathematicians. It consists in systematically replacing ordinary sets with fuzzy sets – or, equivalently, in substituting the relation of graded membership for that of ordinary set membership. This fuzzification program (Zadeh 1965) has produced a large number of more or less predictable generalizations, to the point of having been dubbed a growth industry (Arbib 1977).

(The concept of a fuzzy set is usually regarded as a generalization of that of ordinary set. However, it can be shown to be reducible to that of a family of ordinary sets. Recall how the concept of graded membership is introduced. One starts off with the characteristic function χ_A of a subset A of a given universe U:
$\chi_A: U \rightarrow \{0,1\}$, such that $\chi_A(x) = 1$ iff $x \in A$, and $\chi_A(x) = 0$ iff $x \notin A$.
One then generalizes the range of the characteristic function to the entire real interval $[0,1]$, or some other structure, and introduces the notation

$$x \in_v A \text{ iff } \chi_A(x) = v, \text{ where } v \in [0,1].$$

The *LHS* is read "the degree of membership of x in A is v". This degree is left undefined: it is simply assigned on some extramathematical ground. However, this fuzziness can be removed as follows. Take the power set 2^A of A – where A is again a subset of a given universe U – and ordinary membership in every member of that family of sets. Moreover, assign every subset V of A a certain measure $\mu(V)$ normalized to unity. The concept of graded membership can then be defined in terms of the concepts of ordinary membership and measure, namely thus:

$$x \in_v A =_{df} [\chi_A(x) = v, \text{ with } v = \mu(V), \text{ iff } x \in V \in 2^A].$$

Clearly, this definition holds only for μ-measurable sets. In this case there are as many membership degrees as there are sets, with different measures, in 2^A. In particular, $v = 1$ for $V = A$, and $v = 0$ for $V = \varnothing$. This completes our exactification of membership fuzziness.)

Let us now move on to the typical activity of the mathematician engaged in exomathematical applications, namely *model building*. To begin with let us consider mathematical models in philosophy. There are none in logic *stricto sensu* because the logical theories are maximally general, whereas models are specific. Philosophical models are similar to models everywhere else in being specific theories. In fact a philosophical model is a mathematical theory about a circumscribed class of philosophical concepts. Examples: our own theories of reference, sense, and meaning (Vols. 1 and

2), and of systems, spacetime, and mind (Vols. 3 and 4). On the other hand most of our views in Vols. 5 and 6 are at best condensation nuclei for mathematical models. (And, of course, what philosophers call the 'covering law model' is a thesis, and 'Nagel's model of reduction' is a definition.)

A possible formalization of the notion of a philosophical model is embodied in

DEFINITION 1. A *philosophical model* is a theory about a triple

$$\mathcal{M} = \langle D, F, I \rangle,$$

where

(i) D, the *domain*, or reference class of $\mathcal{M}$, is a a set of concepts or entities of philosophical interest;

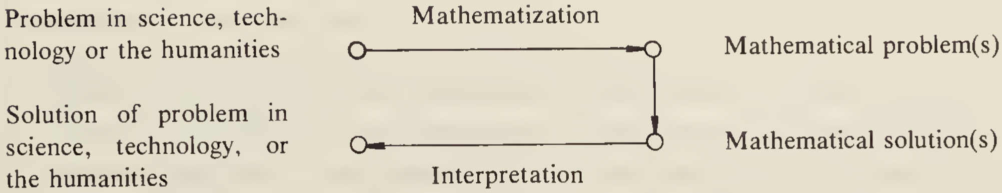
(ii) F, the *formalism* of $\mathcal{M}$, is the union of certain theories in pure mathematics (e.g. predicate logic, group theory, and elementary point set topology);

(iii) I, the *interpretation* of $\mathcal{M}$, is a partial function from the formalism F to the power set of the domain D, that assigns to some predicates and formulas in F sets of concepts or entities in D.

Example 1 The probability theory of truth interprets the probability of a proposition as its degree of truth. (It does not work: see Vol. 2, Ch. 8, Sect. 4.1.) *Example 2* The probability theory of causality interprets the conditional probability of an event given that another event happens, as the strength of the causal link between the two. (It does not work: see Bunge 1982a) *Example 3* Automata theory, enriched with a set of suitable semantic assumptions, is a simple (hence highly idealized) model of things that are at the mercy of their environment (Bunge 1973b, Ch. 8, Sect. 5).

Like art, mathematics can be *figurative* (or representational) or *non-figurative* (or "abstract"). Thus, whereas the models in our semantics are nonfigurative, those in our ontology are figurative (though symbolic rather than iconic) in the sense that they refer to concrete entities. All of the mathematical models in factual science and technology are likewise figurative in this sense – though they are rarely pictorial. This is one of the advantages of modern man over his predecessors: that he can approach reality equipped with a huge tool box full of general (hence portable) ideas, mainly those of mathematics. Hence the difference in outcome, namely mathematical models in addition to pictures and verbal descriptions. (Imagine, if you can, what the dwellers of the caves of Altamira or Lascaux would have painted, had they bred bisons instead of hunting them, and built mathematical models of bison farms.) In short, we can approach the study and control of reality with a rich fund of formal ideas, i.e. of ideas that do not describe reality but provide the necessary framework for models that do describe concrete things.

Every applied mathematician is familiar with this flow diagram or some modified version of it:


Problem in science, technology or the humanities
Mathematization
Mathematical problem(s)
Solution of problem in science, technology, or the humanities
Mathematical solution(s)
Interpretation

The diagram is obvious enough but it poses the philosophical problems of elucidating the notions of mathematization and interpretation. The latter is a semantic concept that was tackled in Vol. 2, Ch. 6, Sect. 3. As for the concept of mathematization, it may be construed as the ordered couple composed of deinterpretation (gutting of factual content) and exactification with the help of some mathematical tools. The actual mathematization and interpretation processes may be quite involved. To begin with, mathematization is never unique, although it may look to be when we have only a few tools in our mathematical kit. As for the interpretation difficulties, suffice it to recall the controversies over the mass concept in the 18th century, the energy concept in the next, and the state function in quantum theory in our own days. But the problems posed by mathematization and interpretation can be revealed at a much more elementary level, as shown in the following examples.

Example 1 It is often stated either that it makes no sense to add three peaches to four quince, or that it does make sense, the result being exactly three peaches plus four quince. This pseudoproblem is dissolved by recalling that peaches and quince happen to be fruits, so that the problem makes perfect sense and has a nontrivial solution, namely seven pieces of fruit. (To indulge in pedantry, call $p \subseteq P$ a set of peaches, and $q \subseteq Q$ a set of quince. Further, call $f \subseteq F$, where $F \supseteq P \cup Q$ is a set of fruits including all peaches and quince. Then set $f = p \cup q$, whence $|f| = 7$ since by hypothesis $|p| = 3$, $|q| = 4$.) *Example 2* At first blush combinatorics is a sort of general physics (or ontology) for dealing with all the possible arrangements or configurations of sets of objects of an arbitrary nature. For example it tells us that, no matter what the nature of three objects a, b and c may be, there are just $3! = 6$ ways of arranging them linearly to produce the chains *abc*, *acb*, *bac*, *bca*, *cab*, and *cba*. However, these are just logical or conceptual possibilities, not physical ones. In fact it may well be that the given objects cannot assemble in any way, or only in one way. That is, the number of nomologically possible arrangements need not be 6 but, depending on the nature of the objects, some integer lying between 0 and 6: this number will be determined by the interactions among the components as well as among these and their environment. *Example 3* Consider the usual mathematization of Zeno's paradox of the race course: it is the infinite series with general term $(1/2)^n$. The series adds up to 1, so there is no paradox if this series is regarded as just a mathematical analysis (one out of infinitely many) of the unit distance. The paradox arises only because Zeno is an operationist: he insists on interpreting every one of the infinitely many terms of the series as an action (running). There are similar paradoxes in quantum mechanics and they are dissolved by avoiding operationism: see Ch. 2, Sect. 6.1.

A while ago we clarified the notion of a philosophical model. Let us now elucidate that of a factual model, or model of things or processes (natural or artificial) of some kind:

DEFINITION 1.2 A *factual model* is a theory about a triple

$$\mathcal{M} = \langle D, F, I \rangle,$$

where

(i) D, the *domain* or reference class of $\mathcal{M}$, is a set of factual items, i.e. concrete things or processes in them;

(ii) F, the *formalism* of $\mathcal{M}$, is the union of some theories in pure mathematics;

(iii) I, the *interpretation* of $\mathcal{M}$, is a partial function from the formalism F to the power set of the domain D, that assigns some predicates and formulas in F sets of factual items in D.

It is convenient to distinguish two kinds of factual model: scientific and technological, according to whether it represents naturally existing things or processes, or man-made ones. (A model of a chemical reaction is of the first kind, whereas a production model of a firm is of the second.) So, we make

DEFINITION 1.3 A factual model about a triple $\mathcal{M} = \langle D, F, I \rangle$

(i) is a *scientific model* iff all the elements of D are natural or social objects, and it complies with the general requirements of science (Vol. 6, Ch. 14, Sect. 2.1);

(ii) is a *technological model* iff some of the elements of its domain D are concrete artifacts or plans for human action, and it complies with the general requirements of technology (Vol. 6, Ch. 14, Sect. 2.2).

Finally we are ready for

DEFINITION 1.4 At any given time the fund of knowledge of *applied mathematics* is the collection of scientific or technological models extant at that time.

There is often talk of 'the application of mathematics to reality' and, since Leibniz, some scholars have marveled at the apparently perfect fit of mathematics to reality. Actually there is no such application of mathematics to *reality*: all there is, is a mathematization of some of our *ideas* about reality. More precisely, when building a mathematical model of some concrete thing, be it electron or industrial plant, one builds or borrows a mathematical formalism, and interprets some of its concepts in factual terms. (Recall Vol. 2, Ch. 6.) By itself, the mathematical formalism is ontologically neutral, this being why we can use it again and again in different research fields.

The ontological commitment of a mathematical model of a piece of reality lies not in its mathematical formalism but in the semantic assumptions that specify the kind of thing that the model refers to and what properties of such things some of the concepts in the model represent. So much so that one and the same mathematical formalism may be attached alternative sets of semantic assumptions. (For example, a probability may be interpreted as the strength of the possibility of an atomic transition in one model, as that of a genic mutation in another, and as that of an error in the transmission of a message in a third.) The main interest of semantic assumptions to the philosophy of mathematics is that their presence or absence is what distinguishes applied from pure mathematics. (More on semantic assumptions in Vols. 1 and 2.)

A subsidiary interest of the consideration of semantic assumptions as the trademark of applied mathematics is that they solve, or rather dissolve, the problem of "the right" mathematics for dealing with problems of a certain kind, e.g. the description of change. Until recently it was widely believed that there is a special mathematics of change, namely the infinitesimal calculus. In recent times it has been claimed that category theory, with its focus on arrows (morphisms) rather than on sets, is the basic tool for the description of change. Both opinions are wrong, as shown by the following example. According to Archimedes' principle in hydrostatics, a balloon will move downwards only if its weight exceeds the buoyancy exerted by the air. Here we are describing (and even explaining) change without the help of either the calculus or category theory. There is no such thing as the mathematics of change as opposed to that of rest: there are only factual theories, some dynamical, others static, that include mathematical formalisms. (Moreover sometimes we know enough mathematics to build alternative but empirically equivalent models of a given process.) Change, or its absence, is described by a mathematical formalism enriched with semantic assumptions. In the case of Archimedes' principle these are: (*a*) The body of weight W is in equilibrium with the surrounding fluid iff $W = B$, where B is the buoyancy exerted by the fluid on the body; (*b*) the body moves upwards iff $W < B$, and downwards iff $W > B$.

As for the alleged perfect fit of mathematics to reality, Platonists have no difficulty in accounting for it: they state that every "form" (or at least every "significant mathematical formula") *must* have a counterpart in the real world, even if we have yet to find it. (See e.g. Thom 1975.) But this is just a pious hope. Moreover it explains a nonfact, for every mathematical model is defective in some way or other: it is at the very best a close approximation rather than perfectly true, for it involves some idealization, such as the neglect of some variables representing properties of the thing concerned or of those in its environment. In sum there is no perfect fit of mathematics to reality.

A related but this time genuine problem is Wigner's (1960): How is it possible for concepts without a definite factual reference (the psychologist's "intervening variables") to be necessary in theories that describe satisfactorily scientific observations? This problem is not solved by assuming that there is no difference between pure and applied mathematics, or that there is a pre-established harmony between the mind and the external world. Rather, it is solved by recalling the following points. Firstly, when a mathematical theory is summoned to help solve a scientific or technological problem, only a small portion of it is used explicitly: the rest remains unused as the support of the part that is being explicitly utilized. For example, in solving an ordinary differential equation one assumes tacitly the theorems that guarantee the existence and uniqueness of its solution. (Analog: When driving a car we pay attention to only a few inputs and outputs.) Secondly, every representation of a chunk of reality – be it a mathematical model or a painting – is symbolic and, as such, it has some conventional components. (For example, the choice of coordinate system and of units is conventional. For conventional elements in the plastic arts see Gombrich 1961.) Thirdly, although there are individual differences among model builders, they all share a basic common neuronal organization and are steeped in roughly the same cultural tradition.

Another question asked, particularly by practically-minded people, is this: What makes a mathematical construct applicable? For instance, why have only a few mathematical geometries found application in physics? It won't do to reply that certain constructs just happen to be inapplicable. If there are indeed any such constructs then we still wish to know why they are inapplicable – in view of the Platonic, empiricist, and pragmatist claim that there can be no such constructs. The truth is that "We don't know what will be useful (or even essential) until we have used it. We can't rely upon the concepts and techniques which have been applied in the past, unless we want to rule out the possibility of significant innovation" (Browder and MacLane 1978 p. 348). All we know is that a person may use mathematical tool *X*, sooner or later, only if *X* happens to be in his tool kit or he has heard of *X*.

Mathematics is necessary but insufficient to build mathematical models of some cognitive or practical value in science or technology. In addition, some substantive knowledge and some flair (intuition or "nose") are needed. Otherwise the models will be far off the mark: they will be just mathematical toys. Unfortunately there is a certain tendency among applied mathematicians to play mathematical games, i.e. to investigate neat academic problems instead of grappling with the complexities and dirt of the real world. That tendency is particularly obnoxious in the social sciences, where many a mathematical model is based on more or less plausible (commonsensical) but entirely arbitrary assumptions that entail "precisely stated but irrelevant theoretical conclusions" (Leontief 1982). The net result is that many of the scientists interested in understanding the object of their study are turned away from mathematics.

It has been rightly said that mathematics is a good servant but a bad master. That is, in factual science and technology mathematics should be handled as an instrument to build realistic models solving genuine problems. (This is the goal. Initially it may be necessary to start with brutal simplifications, such as one-dimensional or static models of things.) If the mathematician wants to enjoy utter freedom, in particular freedom from empirical constraints, he must turn to pure research. Here he can invent or generalize almost any construct. In applied mathematics, on the other hand, "the emphasis is on formulae that can be used to answer specific problems rather than on proofs of theorems under conditions of the utmost generality" (Cox 1962 p. ix).

The difference between pure and applied mathematics does not prevent their practitioners from interacting vigorously. It is a matter of historical record that S & T have been a source of mathematical problems, and pure mathematics a store of ready-made conceptual tools that scientists and technologists have always felt free to use. Sometimes mathematical models of real things are built by using well known mathematical tools, at other times the latter have to be wrought *ad hoc*. And occasionally one may start by building a mathematical formalism, only subsequently looking around for applications. Thus one of the founding fathers of quantum theory stated that a good deal of his research work consisted in "simply examining mathematical quantities of a kind that physicists use and trying to fit them together in an interesting way regardless of any application that the work may have. It is simply a search for pretty mathematics. It may turn out later that the work does have an application. Then one has had good luck" (Dirac 1982 p. 603). The mission-oriented mathematician is not allowed such leisurely exploration of ideas; consequently he cannot expect any such "lucky" strikes.

To conclude. Pure mathematics has no ontological commitment: it does not refer to reality. (If it did then it would constitute the a priori science of the world.) However, it is the basic language of S & T as well as its conceptual sharpener and deductive machinery. In other words, mathematics constitutes no factual knowledge but it is an indispensable means for acquiring precise and profound factual knowledge. This conception of the nature and role of mathematics may be called *instrumentalist formalism*, or *formal instrumentalism*. It differs from the instrumentalist (or pragmatist) epistemology in that (*a*) it does not state that mathematical formulas are rules or instructions rather than propositions, (*b*) it does not make practice the value criterion, and consequently (*c*) it does not reject the mathematical ideas that have not yet found application, any more than those that are no longer widely used in S & T.

4.2. *An Example: Probability*

No concepts have been greater sources of inspiration and trouble, for mathematicians and philosophers alike, than those of infinity, space, and probability. Fortunately most of the fog surrounding these concepts has been dissipated in the course of the last hundred years or so. Cantor taught us that (actual) infinity is an entire (infinite) family of concepts forming the so-called hierarchy of transfinite numbers. Geometers and topologists have created uncounted spaces, and Einstein taught us to distinguish these constructs from (the unique) physical space, and correspondingly the mathematical geometries from physical geometry. (Recall Vol. 3, Ch. 6 and see

Torretti 1978.) The problem of probability differs from that of space in that there is essentially a single core mathematical concept of probability (though the general theory defines an infinite class of probability measures). The analogy between space and probability lies in that both can be variously interpreted. In fact there are at least five different probability concepts counting the one elucidated by the pure or basic theory of probability. Let us sketch them following Bunge (1981b).

Probability theory is a branch of pure mathematics and, more particularly, a chapter of measure theory. In fact a probability measure is a real valued and bounded function P defined on a family A of *abstract* sets. The only conditions that A must satisfy – i.e. the ones that guarantee its (formal) existence – are the following purely formal and rather mild ones: (*a*) the complement of every member of A is in A; (*b*) the intersection of any two members of A are in A, and (*c*) the countable unions of any elements of A are in A. (In sum, A must be a sigma algebra.) The function P is specified by two or more axioms, depending on the theory. (See Fine 1973 for a review.) For example, in Kolmogoroff's theory two axioms suffice to define (implicitly) the concept of absolute probability, whereas in Renyi's three axioms are required to define the concept of conditional probability.

Nothing is said in these theories about events (except in the Pickwickian sense of being members of the family A) or their frequencies, let alone about methods for calculating or measuring probabilities. These other notions occur in the applications. The various probability theories are semiabstract, in the sense that (*a*) they do not specify the nature of the elements of the probability space A, whereas (*b*) they do specify the range of the probability measure, which is the unit interval of the real line rather than another abstract set.

Were it not for this partial semantic indeterminacy, the probability calculus could not be applied almost everywhere, from physical science and physiotechnology to social science and sociotechnology. As long as the probability space A is left uninterpreted, probability has nothing to do with anything extramathematical: $P(x)$, where x belongs to A, is just a number. This number is interpreted as the strength of the possibility of x only upon interpreting A as a collection of facts. Therefore any attempts to define the general concept of probability in specific terms, such as those of favorable case, or degree of belief, or relative frequency in a sequence of trials, is bound to fail. These concepts did have some heuristic value in the past, and they occur in some applications, but they have no place in the basic theories of probability, which are maximally general. Having secured the general and rigorous concept we may hope to apply it to a wide range of problems.

An *application* of probability theory is a factual model of the kind discussed in Sect. 4.1: it is a probabilistic model of the factual items concerned. More precisely, we make

DEFINITION 1.5 A factual model about a triple $\mathcal{M} = \langle D, F, I \rangle$ is a *probabilistic model* of the domain D of factual items iff

(i) the formalism of $\mathcal{M}$ contains a probability calculus or part of it, and

(ii) the interpretation I of $\mathcal{M}$ specifies (a) that the domains of the probability measures occurring in F are included in the domain D of factual items (states of things or events in things), and (b) the probability $P(x)$ of every factual item x measures the chance, or degree of possibility, or propensity of x.

DEFINITION 1.6 A collection D of factual items is called *probabilistic* (or *stochastic* or *random*) iff there is a (sufficiently) true probabilistic model of D (so that every element of D is assigned a definite probability) that is nontrivial (i.e. such that there is at least one probability value different from either 0 or 1).

Contemporary science and technology are chock full of probabilistic models, many of which are true to fact at least to a first approximation. In other words, science and technology acknowledge the existence of a large number of probabilistic (or stochastic or random) collections of factual items, i.e. of things that behave probabilistically. Suffice it to mention but a few well known probabilistic models: of coin flipping and of lotteries; of elementary "particles", atomic nuclei, atoms, molecules, and crystals; of genic mutation and of evolution; of the transmission of signals along nerve fibres or information networks; of learning and of social mobility.

Our Definitions 1.5 and 1.6 contain a covert assumption: they assume that *the* correct interpretation of a probability space A is that of a collection of (random) *factual items*, and that of $P(x)$, for every x in A, is that it quantitates the *objective possibility* of x. (To put it negatively: neither A nor P have anything to do with subjectivity – except of course that both are assumed or checked by some knowing subject or other.) In other words, the two definitions summarize the *objectivist* or *realistic* interpretation of probability. This view is of course in line with any realistic epistemology (Vols. 5 and 6), but at variance with alternative philosophies of knowledge. Let us justify that interpretation and criticize its rivals.

The realistic or objectivist interpretation of probability can be found in

Poincaré (1903), Smoluchowski (1918), Fréchet (1946) and a few other scientists. (In particular, Fréchet regarded probabilities as physical properties on the same footing as distances or masses.) This interpretation was reintroduced independently by Bunge (1951) and by Popper (1957), and adopted, in some guise or other, by several other philosophers (see Settle, 1974). My own favorite arguments in support of it are as follows.

Firstly, when wishing to make sure that choice or preference will play no role in making some decision, we flip a coin or use a table of random numbers. This procedure presupposes the hypothesis that chance is objective and, more precisely, a property of every one of the individual items forming the basic probability space A.

Secondly, in many cases probabilities can be measured, just like any other substantial properties, with the help of objective indicators. For example, counting the long run frequency of repetitive events of a given kind one can estimate their probability. And measuring the intensity of a monochromatic light source allows one to estimate the probability of the transitions of its atoms from one state to another.

Thirdly, in many a probabilistic model occurring in science and technology the probability space A can be built out of the state space of things of the kind being studied. For example, one may set A equal to the power set of the state space. In this way every member of the probability space A is a bunch of states, and $P(x)$ becomes the strength of the tendency (or propensity) the thing has to dwell in the state or states x. Similarly, if x and y are states (or sets of states) of a thing, the conditional probability of y given x, i.e. $P(y|x)$ is interpreted as the strength of the tendency or propensity for the thing to go from state(s) x over to state(s) y.

Fourthly, the whole point of any probabilistic model in science or technology is to account for the behavior of objectively real things. So much so that all the probabilities occurring in them are probabilities of states or of changes of state observed or assumed to occur in external things. Thus when a physicist calculates or measures the probability of an atomic collision of a certain kind, he refers to an event in the real world, not to his own state of mind; and he makes an assertion about an individual event, not about a collection of events. Of course eminent physicists will be found to assert that "the probability concept is in a sense subjective" (Feynman *et al.*, 1963, p. 6–7). However, they fail to justify this claim. Moreover, when computing or measuring probabilities they do not regard their own results as items of (or based on) uncertain knowledge. There need be no more uncertainty about a probability than about a duration or a population count.

Let us now examine the main alternative philosophies of probability: the logical, subjectivist (personalist), and empiricist (or frequency) views. According to the former, probability is a certain relation between propositions: it explicates the notion of confirmation of a hypothesis by the empirical evidence relevant to it. This relation is supposed to be logical, hence independent of fact; consequently the theory built around it, namely inductive logic, is pronounced analytic (Carnap 1950) – a tragic dénouement for a story that originated in the empiricist praise of induction as the method of science.

The root trouble with the logical interpretation of probability is that there *are* no objective procedures for assigning probabilities to propositions, in particular to probabilistic hypotheses such as probability distributions. In science and technology such assignments are made either on the strength of measurements or of hypothetical random mechanisms, such as blind shuffling and genic mutation. But of course in these probabilistic models it is facts (states or events), not propositions, that are assigned probabilities. Moreover, there is nothing a priori about such assignments: they are all supposed to be corrigible in the light of new empirical information. In sum, the logical interpretation does not fit science or technology. It is also useless in pure mathematics, where the domain A of a probability measure is a family of abstract sets. To be sure these can be interpreted as propositions, since the latter satisfy formally the laws of the algebra of sets. But by forcing this particular interpretation one loses the great generality (hence portability) of the probability calculus. In short, the logical interpretation of probability fits neither pure nor applied mathematics. It is but a philosophical mirage – and a declining industry.

(It might be thought that the probability of a proposition can be interpreted in turn as the probability of the fact referred to by the proposition, so that the logical view would be suitable for applied mathematics after all. If this were so, why not interpret the arguments of a probability function as facts to begin with? Besides, the interpretation of the probability of a proposition as the probability of the fact represented by the proposition is possible only in the case of individual propositions. If h is a universal hypothesis, then '$\mathrm{P}(h)$' cannot be interpreted in terms of facts because there are no general facts.)

Understandably the vast majority of the applications of probability in philosophy assume the logical interpretation and, particularly, the hypothesis that propositions can be assigned probabilities. Indeed philosophers of the most diverse kinds – such as Reichenbach, Carnap, Popper, Hintikka, and Good, to mention only a few – have attempted to give probabilistic elucidations (usually reductions) of concepts as diverse as those of certainty, credibility, information, truth, content (or lack of it), confirmation, simplicity, and others. Since such theories never get down to numbers, except in the cases of tautologies and contradictions, they accomplish nothing even though they look exact. Another mistake some philosophers have incurred is holding that there is such thing as "probabilistic inference", which would be exactified and systematized by the probability calculus. Actually every valid reasoning with probabilities is an instance of the inference rules in the predicate calculus: the occurrence of probabilities in the premises makes no difference to the way the conclusions follow. In short, the theory of probability is not a generalization of deductive logic.

The *subjectivist* (or personalist) interpretation construes every probability value $P(x)$ as a measure of the strength of someone's belief in x, or as the accuracy or certainty of his information about x. (See de Finetti 1972, Jeffreys 1957, Savage 1972.) This view was historically the first and it is still very popular because it harmonizes with classical determinism, which has become part and parcel of the vulgar world view. In fact if the universe is deemed to be strictly deterministic (in the narrow sense), probability is resorted to only because of our ignorance of details and of the ultimate causes: God would have no use for probability. The paragon was the classical kinetic theory of gases: here the laws are causal, but probability is used allegedly because of our ignorance of the initial positions and velocities of the atoms or molecules. (This is not true: the theory does not say anything about such ignorance. Instead, it contains the hypothesis that the initial positions and velocities are distributed at random.) There are a number of objections to the subjectivist view, any one of which suffices to ruin it.

A first objection, of a mathematical nature, is that the formula "$P(x) = y$" makes no room for a subject u and the circumstances v under which u estimates his degree of belief in x, under v, as y. In other words, the elementary statements of probability theory are of the form "$P(x) = y$", not "$P(x, u, v) = y$". Such additional variables are supernumerary, and yet they would have to be introduced in order to account for the fact that different subjects assign different credibilities to one and the same item, as well as for the fact that one and the same subject is likely to change his beliefs not just in the light of fresh information but also as a result of sheer changes of mood – not to speak of spiritual crises altering a person's whole world view. In sum, the subjectivist or personalist interpretation of probability is adventitious, in the sense that it does not match the structure of the mathematical concept.

Second, there is empirical evidence against the thesis that our beliefs are actually so rational that they satisfy the probability calculus (or any other theory). For example, most of us hold pairs of beliefs that, on closer inspection, turn out to be mutually contradictory, so that they ought to be assigned nil credence. Another example: most of us do not believe in objective chance, in particular in mere coincidence (or statistical independence). Hence we tend to overrate the probability of the conjunction of independent events (Bar Hillel 1979).

(The subjectivist can circumvent this objection by claiming that the "calculus of beliefs" is a normative theory not a descriptive one. He may indeed hold that the theory *defines* the concept of rational belief, so that anyone whose beliefs do not conform to the theory departs from rationality rather than refuting the theory. Or he may claim that probability theory must be construed as a normative theory of rational behavior under risk, i.e. as decision theory (Wald 1950). The first move will save the interpretation from any ugly unfavorable empirical evidence, but by the same token it will deprive the interpretation of empirical support. And the second move fails for the following reasons. First, decision theory uses probabilities and therefore presupposes an independent theory of probability. Second, the probabilities occurring in standard decision theory are subjective, whereas a rational person is supposed to try and act always on objective probabilities. In real life people who maximize their expected utilities using subjective probabilities are said to indulge in wishful thinking instead of engaging in rational behavior. See Ch. 5, Sect. 5.2.)

Third, it is not true that probabilities are always assigned on purely subjective grounds, i.e. on no grounds whatever. In science and technology one assigns probabilities with the help of observation or measurement. (Even the so-called "principle of indifference", or hypothesis of equiprobability, often adopted in the absence of precise information, is not a principle but just a conjecture supposed to pass empirical tests.) For example, genetic models assign definite objective probabilities to certain genic changes, and the experimental biologist is in a position to test such theoretical values by contrasting them with observed frequencies. Of course we often make probability assignments, or conjecture probability distributions, in the face of insufficient information. But we do the same with many other "quantities" as well, i.e. we hypothesize or guesstimate probabilities just as we conjecture lengths or weights. (On the other hand nobody knows how to estimate the probability of either data or hypotheses. We do not even know what it *means* to say that such and such a formula has been assigned this or that probability.) In sum, the subjectivist interpretation of probability is wrong. (See further objections in Marshak *et al.* 1975, Harper and Hooker, Eds., 1976, and Bunge 1976a, 1981*b*.)

(One way of bringing home the difference between objective indeterminacy and subjective uncertainty is to consider the Prisoner's Dilemma. Of three prisoners, named *A*, *B*, and *C*, only one is to be freed but neither of them knows which one it is to be. *A* asks the jailer to tell him whether *B* or *C* is to remain imprisoned. The jailer, believing that he is imparting no information about the chances of *A*, replies truthfully that *B* will stay in prison. If *A* is a subjectivist he will jump for joy, believing that his chances of being released have suddenly risen from $2/3$ to $1/2$ – just because the jailer blabbered. But if he is an objectivist he will realize that there are two possibilities: either all prisoners have already been sentenced, or only *B* has. If the former is the case, then *alea jacta est*: there is uncertainty but no probability. And if the latter is the case, then there are in turn two possibilities: either *A* or *C* will be judged, and one of them sentenced, or their fate will be left to chance – e.g. to the outcome of coin flipping. Being an objectivist, *A* knows that his *objective chances* cannot be influenced by what the jailer told him: only his own *uncertainty* has changed as a result. Moreover he realizes that, if he and *C* are to be judged, *neither can be assigned a probability* of being freed, for there are probabilities only where there is randomness. It is only if the judges decide to break the law, and commit the fate of *A* and *C* to the coin, that a probability will appear. And this probability is objective and it happens to coincide with the degree of certainty, namely $\frac{1}{2}$ – provided the coin is fair and the judges throw it fairly. The moral is clear: indeterminacy differs from uncertainty, so no prisoner should take consolation in personalist probability.)

Empiricists tend to believe that the correct alternative to subjective probability is frequency. There have even been attempts to define probability as a sort of limit of the relative frequency as the population size increases (Venn 1888, von Mises 1972, Reichenbach 1949). These attempts have been shown to be mathematically incorrect (Ville 1939, Fréchet 1946). Yet, if not the definition, at least the interpretation of probabilities as frequencies is widespread among scientists and technologists. My objections to this interpretation are as follows.

Firstly, probability and frequency, though related, are different concepts. For one thing, whereas the former is theoretical, the latter is empirical. For another, they have different mathematical structures. In fact, whereas a probability function is defined on an abstract probability space *A*, a frequency function is defined, for every sampling procedure, on the power set of a finite subset of A, namely the collection of actually observed events (Bunge 1973a, Ch. 4, Sect. 2.5). Consequently a probability statement does not have exactly the same reference class as the corresponding frequency statement. Whereas the former refers usually to an *individual* (though possibly complex) fact, the latter is about a whole set of facts chosen in accordance with certain sampling techniques. (In other words frequencies, like averages and standard deviations, are *collective* properties not individual ones.) For example, one speaks of the frequency (not the probability) with which one's telephone is heard (by someone) to ring over a certain time interval, thus referring to an entire collection of facts.

Secondly, probabilities and frequencies seldom coincide numerically for, whereas the former are independent of the number of trials (or the populations size), frequencies fluctuate with it. Moreover, the two are clearly distinguished in a number of theorems, particularly in the laws of large numbers. For example, in a binary sequence of n independent trials (e.g. coin flippings), the (meta)probability that frequency and probability coincide numerically increases with n. This classical theorem cannot even be *stated* if probabilities are equated (via definition or interpretation) with limiting frequencies. (Actually the frequency occurring in this theorem is a theoretical concept, namely the average probability, which can be construed as the expected frequency.)

Thirdly, whereas every frequency is the frequency of actual observations of facts of some kind, a probability may be interpreted as the quantitation of a potentiality yet to be actualized. (Recall Vol. 3, Ch. 4, Sect. 4.) Consequently equating probabilities with frequencies (via definition or interpretation) involves rejecting real (objective) possibility – i.e. adopting an actualist ontology. (In turn this involves forsaking the understanding of the probabilistic models in science and technology, which take objective potentiality and randomness in earnest.) The equation of probability and frequency involves also restricting the domain of applicability of the probability calculus to observable phenomena, which constitute only a small subset of the collection of facts. Finally, the equation involves employing probabilities only posthumously, after all the observations have been made and sifted – never for predictive purposes.

The correct procedure with regard to the relations between probability and frequency is not to equate them but to clarify their differences and relations. In particular, frequency is among the estimators or indicators of probability. (It is not the only one. Recall that in physics probabilities are often checked or estimated via other "quantities", such as spectral line intensities, temperatures, and scattering cross sections.) In short, there are often frequency *estimates* of probability, but the frequency interpretation of probability is mistaken.

We are left then with only two legitimate probability concepts: those of pure and applied probability, i.e. those of (normed) measure and tendency or propensity, respectively. The other three concepts – the logical, subjectivist, and frequency ones – are illegitimate. The logical interpretation is mathematically correct but far too restrictive and it has no application to science or technology: it is idle. And the other two interpretations are mathematically objectionable, unsuitable for science and technology, and philosophically implausible – at least with respect to the philosophical system that is being expounded in this *Treatise.*

The moral of this story is that pure and applied mathematics are clearly distinct, yet they must not be separated. Each of them has its own goals, which it can attain only by interacting vigorously with the other and by avoiding intercourse with obsolete philosophies.

5. FOUNDATIONS AND PHILOSOPHY

5.1. *Foundation of Mathematics*

Mathematics has foundations, though they are neither permanent nor above doubt. (See Hilbert and Bernays 1968, 1970, Wilder 1952, Beth 1959, Stoll 1963, or Hatcher 1982.) At any given stage in the history of mathematics, its foundations consist in the most general concepts, principles and methods of mathematics: it is the union of logic (*lato sensu*) and metamathematics. To be sure the odd philosopher – e.g. Wittgenstein (1978) or Lakatos (1978) – has held that mathematics stands in no need of foundations. But mathematicians are aware of the existence of foundations if only because the subject occurs together with logic at the beginning of every issue of *Mathematical Reviews.* However, most mathematicians work in areas so far removed from foundations that they need not care about the latter. In fact their work is largely foundation-free, or compatible with alternative foundations. Therefore it is unlikely to be shaken by any revolutions in foundations. Yet it may be considerably enriched, or at least illuminated, by foundational results.

The foundations of mathematics, or of any other science, has several motivations. One of them is *philosophical*: the wish to harmonize foundations with some philosophical system, such as rationalism, intuitionism, or empiricism. Another is *psychological*: the wish to attain ultimate certainty. The logicists sought the latter in the reduction of all mathematical objects to logical ones; the formalists, in proving the consistency of every mathematical theory; the intuitionists, in relying on a handful of allegedly infallible intuitions. All three moves failed, and yet foundational work went on: we have come to realize that "A Foundational system serves not so much to prop up the house of mathematics as to clarify the principles and methods by which the house was built in the first place". (Goldblatt 1979 p. 14). Moreover we have come to expect that the foundations of mathematics will continue to shift and even branch out as a result of research into them. Thus a distinguished foundational worker notes that "the basic theory of elementary toposes [...] seems to be almost completely worked out [... Doubtless there are minor points to be cleared up,] but the foundations of the subject do appear to be pretty stable. This is of course a bad thing: it is vital to the health of a subject as basic as topos theory that its fundamental tenets should be the subject of continual review and improvement, and I am uncomfortably aware that by writing this book I have contributed largely to the concreting-over of these foundations" (Johnston 1977 p. xvi).

A third motivation for foundational work is the yearning for *simplicity*, which would be satisfied by reducing the whole of mathematics to a single theory (e.g. logic or arithmetic). This too has been and continues to be a powerful rationale even though it is but a mirage, as shown by the progressive complication of the field. Finally, a fourth motivating force is the wish to *systematize* the various chapters of mathematics: to put them together into a single vast system. This is perhaps the most legitimate of the four motivations, because it is realizable and it has yielded abundant fruit, such as the formal systems underlying their various models, and the functors relating the various systems. Besides, it has the advantage of promoting cooperation among specialists and facilitating the understanding of mathematics. In sum, there is no question that mathematics has foundations, albeit shifting ones that are unpopular subjects of study even among mathematicians.

Those few mathematicians who do care about foundations are either satisfied or dissatisfied with standard logic and metamathematics. The former, who constitute the vast majority, believe that standard (ordinary, "classical") mathematics is basically correct and has never been so prosperous. Their critics worry about the standard logical foundations of mathematics, to the point of speaking occasionally of "a crisis in contemporary mathematics [...] *due to our neglect of philosophical issues*" (Bishop 1975 p. 507). Now, it is true that the foundations and the philosophy of mathematics are intimately related. (The same holds for any other science.) In fact the type of foundations one favors is usually motivated and justified by philosophical theses concerning the nature of mathematical objects and the way we get to know them. For example, logicism has been wed to Platonism, formalism to nominalism, and constructivism to intuitionism.

However, it is possible and convenient to separate foundational from philosophical issues, in view of the fact that the above mentioned historical associations are causal rather than logical, as shown by their considerable weakening in recent years. Thus it is possible to adopt intuitionistic logic without admitting its original philosophical motivations or even without adopting constructivism. And, although constructivism was initially associated with Kantian intuitionism, and even with operationism (see Bunge 1962a), most contemporary constructivists seem to be indifferent to philosophical matters. They do not seem to believe that "intuitionistic mathematics is pointless without the philosophical motivation underlying it" (Dummett 1977 p. viii) – which is fortunate because that motivation is controversial to say the least. For those mathematicians constructivism is an interesting research strategy free from the risks of alternative policies. Many constructivists are just curious to find out how much "classical" mathematics can be rebuilt in a constructive fashion ("constructivized"), and what new vistas the constructivist approach may open. Likewise one may adopt a logicist strategy, or a formalist one, or even a combination of the two, without admitting any of the philosophies traditionally linked to them. In short here, as in every other research field, current work can be dissociated from its historical roots. We may then, up to a certain point, and breaking the tradition, uncouple foundational matters from philosophical ones.

Now, there are two classes of foundational strategy or policy: *restrictive* and *non-restrictive*. The latter demand only that logical conditions be obeyed, in particular exactness, systemicity, and non-contradiction: aside from these minimal constraints the mathematician should feel to invent whatever he feels like. This is of course the strategy tacitly adopted by "classical" mathematics and the one explicitly advocated by logicism. The restrictive strategies, on the other hand, set additional constraints on mathematical inventiveness. Thus formalists and constructivists alike reject actual infinity and admit only finitistic methods. The main difference between them, aside from their motivation, is that the constructivist demands in addition that all concepts and proofs be constructive at least hypothetically, i.e. in principle.

We proceed to examine briefly these various foundational strategies without regard for the philosophies supposed to buttress (or discredit) them. (By dismissing any of them on purely philosophical grounds we would risk missing important mathematical contributions.) Nor shall we waste much time on propagandistic claims such as "Policy X is more *natural* than policy Y", or "Concepts of type X are more *meaningful* than concepts of type Y", because their proponents do not bother to tell us what "natural" and "meaningful" mean to them. (In particular, intuitionists are remarkably vague with regard to the words 'intuitive' and 'meaning', which they use very often. Thus Dummett (1977), an eminent philosopher of intuitionistic mathematics, offers no theory of meaning.)

It is well known that the classical schools in the foundations of mathematics are logicism, formalism, and intuitionism. (See Wilder 1952, Beth 1959, Hatcher 1982.) These schools were formed roughly between 1890 and 1910, but they have mellowed and even cross-fertilized since then, to the point where there are hardly any pure logicists, formalists, or intuitionists left. Let us review them quickly.

Logicism, advocated mainly by Frege and Russell, and favored by most logicians and exact philosophers, comes in two strengths. The *weak* thesis is that every mathematical theory contains logic – or, equivalently, that every mathematical theory has a logical foundation. (I.e., if $\mathscr{T} = \langle F, \vdash \rangle$ is a mathematical theory, the set F of its formulas contains the totality of tautologies of some system of logic.) This thesis is controversial except to Brouwer and his faithful, who claim that mathematics precedes logic. It is also philosophically rather uninteresting.

The *strong* logicist thesis states not only that mathematics is based on (i.e. presupposes or contains) logic, but also that it is fully *reducible* to it: that every mathematical concept is definable in logical terms, and every mathematical formula is derivable from logical formulas. (Recall the definition of full or strong theory reduction: Vol. 6, Ch. 10, Sect. 3.1.) However, this thesis can be construed in two different ways, according as logic is conceived in a strict (narrow) or in a wide sense. According to the former construal, mathematics is reducible to elementary logic, i.e. to the first order predicate calculus. Clearly, this version of logicism is false, as shown by the following two elementary counterexamples. The abstract relation of partial order is not reducible to (definable in terms of) any of the concepts occurring in first order logic; the same holds for the abstract operation of concatenation, which occurs in algebra. In short, mathematics is not reducible to first order logic; in particular it is not a heap of tautologies.

We must therefore construe the logicist thesis in its lax sense, namely as the contention that mathematics is reducible to logic *lato sensu* – i.e. the predicate calculus of the first and higher orders, model theory, set theory, and possibly further universal theories (such as category theory) as well. If understood in this wide sense, it must be admitted that, like many another reductionist strategy, logicism has had some stunning successes. An early victory was the reduction of arithmetic to the predicate calculus augmented with set theory. But, like every other radical reductionist program, logicism has been shown to have limitations. For one thing, since there are alternative set theories (every one of which defines implicitly its own membership relation), there can be no *unique* reconstruction of mathematics on the basic of logic *lato sensu*. So, the original program of logicism should be reformulated in a more realistic way, namely this: "Reduce as far as possible but recognize the limits to reduction when you encounter them. When you do encounter such limits you are facing new, emergent constructs, related to but not reducible to the previously introduced ones". Even cut to this size, logicism is invaluable for its insistence on exactness and on the need for laying bare the logical basis and the logical structure of mathematical theories.

Logicism has often been misunderstood by philosophical friends and foes. One such misunderstanding is equating it with the thesis that, "in principle", mathematicians need know only the predicate calculus and set theory to work, say, in number theory. Actually, even when doing the simplest sums we had better forget the reduction to logic, because the bridging or defining formulas (e.g. $\ulcorner 1 =_{df} \{\varnothing\} \urcorner$) can get so cumbersome, that the arithmetical propositions become unmanageably long and unperspicuous. Consequently "we continue to do arithmetic as before only with the awareness that there is a sense in which our proofs and calculations could be translated into set theory" (Wang 1966).

A second popular misunderstanding of logicism is revealed by those critics who hold the logical reconstruction of a piece of mathematics to be circular for presupposing (in the epistemological sense) a prior intuitive knowledge of that piece (Casanova 1947, Steiner 1975). This objection misses the point of any rational reconstruction, which is not to build new theories, let alone to teach them, but to bring order and possibly also consistency into a somewhat messy body of knowledge that has grown naturally in a zig-zagging historical process. Logicism is not to be understood as the thesis that logic oozes mathematics, but that mathematics can be reconstructed out of logical constructs alone. That such reconstruction is not always possible, is another matter.

The next foundational strategy to be examined is *formalism* (Hilbert 1935, Hilbert and Bernays 1968, 1970, Curry 1951, Robinson 1965). Formalism too, has changed over the years, though perhaps to a lesser extent than logicism, and it has been no less misunderstood than logicism. For example, Davis and Hersh (1981), following Lakatos (1976), have attacked formalism for failing to provide a faithful account of the actual process of mathematical research. The latter, as we know, is no less zigzagging and subject to error and uncertainty than any investigation in natural science. But the point of formalism was not to account for the psychology and history of mathematics: it was to reconstruct in a rigorous way the outcome of a rather unruly process. Moreover this clean copy had to meet the objections that Brouwer had raised against the logicist reconstruction, and it had to save the bulk of mathematics, which had been threatened by Brouwer. The rigorous reconstruction was intended to culminate a process of spontaneous growth, not to replace it, let alone to describe it.

The main technical features of formalism are abstraction, finitism, axiomatics, and metamathematics or proof theory. (We regard the nominalist thesis, that the basic mathematical objects are marks or signs – e.g. numerals instead of numbers – as a philosophical thesis. Accordingly we shall deal with it in Sect. 5.2.) Let us begin with abstraction. Although there had been abstract theories before – e.g. Boolean algebra and group theory – Hilbert stressed their importance as well as the radical difference between such theories and those where every symbol is interpreted in a "concrete" (possibly also familiar) fashion. (Recall what was said about model theory in Sect. 3.1.) Hilbert succeeded even in digging up the abstract theory underlying elementary geometry – which won him the hatred of intuitionists.

A second feature of formalism was finitism: the attempt to complete the process of eliminating infinity from mathematics – a process started by Weierstrass when he replaced intuitive considerations about limits, as well as "infinitely small" and "infinitely large" quantities, with reasonings about finite deltas and epsilons. To be sure Hilbert dit not reject actual infinity but regarded it as a fiction, and correspondingly all talk about actual infinity as a *façon de parler* – as if finite sets were any less fictitious. (See Hilbert 1925.)

A third feature of formalism was the program of axiomatizing all mathematical (and also physical) theories. The aim of axiomatics is to seize on the basic ideas (primitive concepts and postulates) and arrange them in an orderly and consistent manner (See Bunge 1973a, Chs. 7 and 8.) This can only be attained by disregarding historical accidents and pedagogical devices. (Inevitably, this feature of formalism was to irritate the lovers of confusion as well as of the people unaccustumed to viewing theories as wholes.) A fourth characteristic of formalism was the systematic metamathematical study of theories and, in particular, the attempt to prove their consistency by finitist methods.

All four proposals came eventually under fire. The high regard for abstraction, joined with the nominalist thesis that the basic objects of mathematics are marks on paper rather than concepts, was taken for the ludic doctrine according to which mathematics is "a sort of game, played with meaningless marks on paper rather like noughts and crosses" (Ramsey 1931 p. 68). Had Hilbert been a consistent nominalist he would have espoused the ludic doctrine, for this follows from identifying the objects of mathematics with symbols, and its laws with conventional rules. Fortunately he was not consistent in this regard, as shown by the kind of problems he grappled with or proposed to others: they were anything but games. Much the same can be said about the early Whitehead, who had defined algebra "as an independent science dealing with the relations of certain marks conditioned by the observance of certain conventional laws" (1898 p. 11). In any case the view of mathematics as a game did not last except among some philosophers such as Wittgenstein (1978). With the rise of semantics people realized that, unlike mathematical theories, games have no metatheories and do not involve a notion of truth (Fraïssé 1982).

The most vulnerable thesis of formalism turned out to be the one about the scope of formalization. It suffered a severe blow when Gödel (1931) proved that every consistent theory containing arithmetic is necessarily incomplete. However, in my view this setback did not dash Hilbert's program of axiomatizing all mathematical theories and proving their consistency. It only *restricted* the scope of axiomatics and killed the hope of proving the consistency of mathematical theories with the sole help of *finitist* methods. The axiomatic format, in particular the axiomatic definition of new concepts, has become standard not only in the foundations of mathematics but in all of abstract mathematics, particularly in the work of the Bourbaki team – which, incidentally, embraced the formalist strategy (Dieudonné 1981) though without caring for foundational studies. Axiomatics has also been used in the foundations of science (see e.g. Bunge 1967c, 1973a and Suppes 1969) as well as in philosophy (see the first four volumes of this *Treatise*). Besides, the line of metamathematical investigation initiated by Hilbert and Peano has proliferated, though with two curious features. The first is that it has flourished outside the main formalist school, the Bourbaki team. The second is that most of its results have been negative, i.e. of the type "Theory x is incomplete", and "Theory y is non-categorical". In short formalism, just like logicism, is far from dead. In a weakened form and dissociated from nominalistic philosophy, formalism has become part of the mainstream of the foundations of mathematics.

We now leave the mainstream, though by no means the vanguard, to have a look at intuitionism. Brouwer, the founder of the school, championed both radical constructivism and Kantian intuitionism, as well as some mystical ideas that have been set aside by his followers. It may be said that his slogan was *No existence without construction, and no construction without intuition.* Brouwer crusaded against "classical" or standard mathematics for being full of non-constructive concepts, such as that of a continuum, and non-constructive proofs (by reduction to absurdity). In particular he did not accept the excluded middle and he rejected actual infinity as well as the axiom of choice. He also denied the priority of logic. And he failed to recognize that constructivity, surveyability and intuitability are not absolute but come in degrees.

Intuitionism has changed considerably since Brouwer's day: it is anything but a small homogeneous sect. First Heyting formalized intuitionistic logic, a feat seen by many as a suicidal concession to formalism. Later on a number of mathematicians embraced constructivism while paying no attention to the philosophical underpinnings of intuitionism. (This convergence was unavoidable given that formalism too is constructivist: see Curry, 1951 and Markow, 1975.) More recently, the computer revolution has reinforced the constructivist camp with a huge crowd of computer scientists convinced that mathematics is a collection of algorithms. And lately a number of mathematicians have tackled abstract mathematics – a contradiction in terms to a true intuitionist – on the basis of intuitionistic logic though without admitting constructivism. So, 'intuitionism' has come to mean different things to different people.

For half a century intuitionistic mathematics yielded only a meagre crop. Consequently adopting it involved sacrificing the bulk of mathematics on the altar of a cruel god fired by a dubious philosophy. Even Heyting (1957) admitted that such sacrifice was "an inevitable consequence of our standpoint". This situation altered radically when Bishop (1967) and a few others succeeded in "constructivizing" a large portion of analysis. (See Stolzenberg 1970 for a paean and Goodman 1983 for a cold shower.) Since then intuitionistic mathematics, or what passes for such, has been thriving. (See Troelstra and Van Dalen, Eds. 1982.)

However, as noted before, not all constructivists are intuitionists: indeed some are formalists. And not all those who call themselves 'intuitionists' are constructivists: some accept intuitionistic logic but make use of nonconstructive principles and methods. Moreover the constructivist principle comes in two strengths: a moderate one and a radical one. *Moderate constructivism* is the thesis that every concept and every proof are to be constructive: that we must always show an explicit routine for constructing our concepts or proofs – hence we must abstain from asserting the existence of anything that we do not know how to construct. For example, we can talk about sets only provided we define them; and we can assert that an equation has a solution only if we know how to find at least an approximate solution to it. In short, *to be is to be constructible.*

Radical constructivism, proclaimed by Brouwer and revived by Bishop, goes much further: it demands that every construction have a "computational meaning" and, in general, that all of mathematics be assigned a "numerical meaning". This means that, however high on the scale of abstraction we may be working, we must always be able to "get down to numbers" – meaning the positive integers. In other words, mathematics should be based on the natural numbers rather than on undefined predicates, abstract sets, or morphisms. *To be is to be a natural number or to be constructible out of natural numbers.*

Radical constructivism has been at work in analysis. Outside analysis, particularly in logic, algebra, and topology, most of the mathematicians who call themselves 'intuitionists' adopt the moderate version of constructivism or just are not constructivists and accept only intuitionistic logic. The reason for this is plain, namely that radical constructivism makes no room for abstract objects, which are the ones occurring in abstract algebra and general topology. (Abstract objects are the *bête noire* of all non-rationalists.)

Since constructivism is gaining ground, we had better take a closer look at it. The hub of constructivism, whether moderate or radical, is the concept of existence. In "classical" (standard) mathematics, exactness and consistency guarantee formal existence: recall Sect. 2.1. In particular, by the completeness theorem of first order standard logic, every consistent theory based on the predicate calculus has at least one model: so, the existence of the objects in any such model is warranted even before we proceed to define them. The constructivist recognizes no such birthright: he admits only existence by construction: he wants to see the baby before christening it. In this he resembles the Greek geometers, who demanded that every geometrical figure and proof use only rule and compass. (This excluded, of course, any reasonings on infinite lines or planes.) He also resembles the operationist, who claims that to exist is to be measurable. A few elementary examples will suffice to appreciate how straight and narrow the constructivist path is.

Example 1 In standard set theory, and in the set-theoretic reconstruction of arithmetic, the most important set is the empty set. Constructivism declares this set to be nonexistent because one cannot construct nothingness. (Remember the times when people doubted that 0, or even 1, were numbers?) *Example 2* In standard mathematics one may prove a theorem (e.g. the assertion that there is a single empty set) by showing that its denial leads to contradiction. This mode of proof (by *reductio ad absurdum*) is not available to the constructivist, who must prove either a proposition or its negation. (Moreover constructivists try to avoid negation: they are literally positivists.) *Example 3* In "classical" analysis one proves the so-called intermediate value theorem: A continuous real function f defined on a closed interval $[a,b]$, such that f is negative at a and positive at b, vanishes somewhere in between. I.e. *there is* a point $a < \xi < b$ for which $f(\xi) = 0$. Constructive analysis does not contain this theorem for it does not tell us how to compute ξ. (Recall Fig. 1.4 in Sect. 3.2.)

This last example is quite typical of the mutilating effects of the constructivist strategy. Not only most analysts but also all factual scientists are keen on the intermediate value property because they use it daily, albeit tacitly most of the time. In fact to go from one place to another we must pass through every intermediate point along our route – and we do so even when we do not know what points we are passing. (This refutes the contention that "the only reason mathematics is applicable, is because of its inherent constructive content" (Bishop, 1975).)

Nor is that the only sacrifice we are asked to make to comply with intuitionism. All the other existence formulas, among them the axiom of choice, are to be given up as well unless they can be coupled to constructive procedures (algorithms). This is not all: orthodox intuitionists must also believe Brouwer's proof that all real functions are continuous. That is, they cannot admit discontinuous functions such as the step function (unless approximated by a continuous function). Hence intuitionism cannot analyze such discontinuous physical processes as the collision of elastic bodies or the refraction of light waves. It is not true, then, that the whole of "classical" analysis has been, or even can eventually be, "constructivized". And, because of some of its limitations, it will never be able to be as useful to factual science and technology as standard analysis. Which is not surprising taking into account that intuitionism is the companion of operationism.

The limitations of constructivism do not entail that it is an unacceptable strategy for theory construction. It *is* admissible provided its imperial wings be clipped. More precisely, constructivism is legitimate provided it tolerates nonconstructive concepts and methods side by side. In other words, we should acknowledge *two methods* of bringing mathematical constructs into (formal or conceptual) existence: postulation and construction. The former, or "classical", method demands only consistency. (If P is a predicate defined implicitly by a system of postulates, then P is non-contradictory iff the extension of P is non-empty, i.e. if there are objects with property P. In this case P may be said to exist conceptually *by fiat*. The scope of this method is somewhat limited because of Gödel's incompleteness theorem.) The constructive method is far more demanding and, therefore, psychologically reassuring: "The only way to show that an object exists [constructively!] is to give a finite routine for finding it" (Bishop 1967 p. 8). There is no justification, other than philosophical bias, for excluding either of these methods. We should make full use of the unique freedom we enjoy in mathematics, letting the factual scientists and technologies worry about the constraints inherent in the real world.

Now a word on an extreme version of constructivism, namely *recursivism* (e.g. Grzegorczyk 1959). The slogan of recursivism is that analysis should boil down to the theory of recursive functions. (A recursive function of one variable takes natural numbers into natural numbers.) Recursion theory was initially invented to exactify the rather hazy notion of computability, and most mathematicians regard it as a smallish chapter of analysis. (It is sometimes included in logic because of its importance in metamathematics, which is like regarding analysis as a branch of physics because of its central role in this science.) The current vogue of the theory is due to its importance in computer science. If computers could have a philosophy they would be recursivists and nominalists.

Recursivism demands an intolerable mutilation of analysis. In fact the most general and interesting functions are the real functions, and particularly the continuous ones, which lie beyond the scope of recursion theory. The latter analyzes the way we *compute* values of functions to any desired degree of accuracy, but it cannot even *define* such elementary functions as the linear and the logarithmic functions. Hence recursion theory may be regarded as ancillary to "classical" analysis, never as a substitute for it. Therefore recursivism, unlike the less radical versions of constructivism, is not a possible foundational strategy. Its inability to recapture a substantial part of "classical" analysis should be taken as a warning against the attempt to reduce all mathematical objects to the positive integers.

In sum, there are alternative foundational strategies which deserve being tried out. Most contemporary mathematicians have tacitly adopted a combination of logicism and formalism. And, although most mathematicians do not care for the strictures of intuitionism, they are all happy when a "classical" result is reproduced by constructive methods. So, all three major foundational strategies have their place in contemporary mathematics, and neither dominates absolutely. For example, new concepts are sometimes introduced constructively in a step by step fashion, as demanded by both logicists and constructivists; and other concepts are introduced by postulate systems (i.e. creatively), as recommended by the formalists. And, whereas some proofs are direct, most are indirect (i.e. they use the excluded middle). middle).

The adoption of a single foundational strategy at this time would impoverish pure mathematics and, consequently, it would diminish its power as a tool for building factual theories. Foundational monism would also jeopardize our understanding of the nature of mathematics, for each of the major foundational strategies does capture one aspect of mathematics. Thus logicists and formalists stress the deductive side of mathematics, whereas intuitionists stress its nondeductive side. Indeed, where the former say that every theorem follows from a set of postulates and definitions, intuitionists claim that it is the other way round: that we fashion the postulates and definitions so as to get the results we want or need. Actually there is no incompatibility between the two theses: the former concerns finished mathematical theories, whereas the latter is about mathematics in the making. (In other words, whereas logicists and formalists focus on the deductive structure of mathematics, intuitionists stress its heuristics.) Hence the two views on the nature of mathematics are mutually complementary rather than incompatible. Adopting one of them to the exclusion of the other is therefore as unwise as opting between biological systematics and evolutionary biology, or between sociology and history. Every major foundational strategy has its merits and shortcomings, and every one of them sheds some light on certain aspects and chapters of mathematics while leaving others in the dark. Consequently we should advocate *foundational pluralism* at least for the time being.

The thesis that alternative foundational strategies are admissible generalizes the 19th century finding that there are alternative geometries, as well as the more recent finding that there are alternative set theories (every one of which exactifies a different notion of membership). This does not entail that all foundational strategies are worth the same, so that the choice among them is purely a matter of taste. As a matter of fact some strategies yield more general and deeper foundations than others. Thus the reduction of mathematics to category theory results in a deeper, more vast building than the reduction to (a) set theory, which in turn goes wider and deeper than the reduction to arithmetic. So, if the desideratum is generality and depth, we know which strategy to choose. But one may have alternative desiderata. The choice of a foundational strategy resembles that of an architectural style. It is partly a matter of function: what do we want the building (or the theory) for? It is also partly a matter of resources: what materials and procedures (assumptions and methods) are we willing to employ? Hence it is also partly a matter of budget: what price are we prepared to pay? Finally it is also partly a matter of taste (philosophy). There is nothing wrong with admitting a plurality of styles in either mathematics or buildings, provided we do not mix them in one and the same construction.

As a matter of historical record each foundational strategy has been partly motivated by a given philosophy: logicism by idealism (in particular Platonism), formalism by nominalism, and mathematical intuitionism by philosophical intuitionism. To be sure, neither has been totally faithful to its sources. In particular the mathematical intuitionist's intuition is supposed to be the basis for a laborious intellectual construction, whereas Kant's pure intuition was supposed to deliver its object directly. Still, the ties are there.

Therefore, if we wish to uphold a pluralistic foundational policy ("philosophy") we must cut the traditional ties of foundations with philosophy: we must build a new philosophy of mathematics compatible with foundational pluralism. This we have done in outline in the previous sections. Let us now complete that sketch and compare it with alternative philosophies of mathematics.

5.2. *Philosophies of Mathematics*

A philosophy of mathematics should propose well-founded answers to such questions as: What is mathematics and how does it differ from the other sciences?, What is the nature of mathematical objects and how do they differ from material objects?, How do mathematical objects exist?, Does mathematics have any ontological presuppositions?, Is mathematics a priori, a posteriori, or both? What is mathematical truth?, What is mathematical proof?, How does mathematics relate to elementary logic and to semantics?, and How can mathematics, which is intemporal, cope with reality, which is changing? There are, of course, many other problems that a philosophy of mathematics should be able to tackle. On the other hand a philosophy of mathematics need not deal with empirical problems that are best approached by factual disciplines such as the psychology and the sociology of mathematics – e.g. How do we fashion or learn mathematical ideas, and How do mathematical communities emerge, evolve, and decay? (Typically, empiricists, intuitionists and dialectical materialists refuse to draw this distinction between conceptual and empirical questions about mathematics. This distinction is necessary, even though a suitable understanding of mathematics as a whole calls for both conceptual and empirical inquiries. We must discern between distinction and separation, and insist on the former while rejecting the latter.)

A fair number of philosophies of mathematics have evolved over twenty five centuries, and nearly all of them are still alive. (See Brunschvicg 1929, and Benacerraf and Putnam Eds. 1964.) This is due not only to dogmatism but also to the fact that every one of them accounts satisfactorily for *some* aspect or other of mathematics while leaving the other aspects in the dark. This can be understood upon reflecting that any philosophy of mathematics may have, but as a matter of fact seldom has, all of the following components:

(i) *ontology*: questions about the ontological status of mathematical objects;

(ii) *semantics*: matters of sense, reference, and truth in mathematics;

(iii) *epistemology*: questions about the nature and sources of mathematical knowledge;

(iv) *methodology*: matters of justification (in particular proof) and application.

Nearly every philosophy of mathematics has focused on a couple of these components, presumably those where it had a positive contribution to make. The remaining components were neglected, either because overlooked or because they proved embarrassing to the general philosophy of which the given philosophy of mathematics was but a branch or an application. In contradistinction to such partial philosophies of mathematics, we need one capable of accounting for all four components above, as well as being compatible with the general philosophical ideas formulated in the previous volumes of this Treatise. But, since we have granted beforehand that a number of philosophies of mathematics have grasped correctly some traits of their object, let us review them quickly and then evaluate them.

The major classical philosophies of mathematics are Platonism (a variety of objective idealism), nominalism (a variety of materialism), intuitionism (a variety of subjective idealism), and empiricism (in particular pragmatism). Their main theses are, schematically, as follows.

PLATONISM

Ontology: All ideas are self-existing: we only grasp some of them (imperfectly).

Semantics: Every non-contradictory construct is meaningful; truth, whether formal or factual, is correspondence with objective (ideal or material) reality.

Epistemology: The soul may remember some of the ideas it was originally exposed to; in any case we recall or discover but cannot invent any ideas.

Methodology: The axioms must be self-evident, and the theorems must follow deductively from them; as for applications, they consist in discovering the mathematical (ideal) properties of things.

NOMINALISM

Ontology: The objects of mathematics are symbols (of nothing), e.g. marks on paper: there are no "abstract entities".

Semantics: Nil.

Epistemology: Doing mathematics is playing a game with symbols according to definite rules. Mathematics has no subject matter.

Methodology: The pinnacle of mathematical research is the construction of formal axiom systems unattached to any particular interpretation. Only deductive proofs in consistent formal theories can be fully rigorous.

INTUITIONISM

Ontology: Mathematical ideas are mental objects and personal creations.

Semantics: Only intuitive mathematical ideas – in particular those which are ultimately reducible to the *Urintuition* of the sequence of natural numbers – are meaningful.

Epistemology: Mathematical knowledge is obtained by intellectual intuition rather than by sense experience or by pure reason. Whatever is counterintuitive (e.g. non-computable numbers and actual infinities) is not really known, hence it is no part of mathematics.

Methodology: Only constructive concepts and proofs are admissible.

EMPIRICISM

Ontology: Mathematical ideas are mental objects.

Semantics: Mathematical ideas mirror or summarize experience.

Epistemology: Mathematical knowledge is gotten in the same way as any other knowledge – in particular inductively. Mathematical axioms are nothing but pills of empirical wisdom.

Methodology: The ultimate test of mathematical ideas is human experience, however indirect – e.g. through the experimental test of scientific theories involving mathematics.

There are of course a few alternative philosophies of mathematics, but they are sketchy and have exerted no influence on mathematical research. The best known are those of Wittgenstein and Lakatos. Wittgenstein (1978) dealt mostly with "household mathematics" (Bernays 1976) – i.e. grammar school mathematics – and, tangentially, with some of the foundational problems investigated by Russell. His central interest with regard to mathematics was its psychology: matters of learning, understanding, inventing, and above all using elementary mathematical ideas. His entire philosophy of mathematics boils down to the pragmatist thesis – already found in Mach – that all mathematical propositions are *rules* (for computing or drawing), so that doing mathematics is a special case of following rules or playing games. This view is utterly inadequate if only because (*a*) it ignores that in mathematics all rules are based on laws, and (*b*) it overlooks the heart of mathematics, which is the building of mathematical systems (e.g. $\langle \mathbb{N}, 0, + \rangle$) and the proving of theorems about them. We shall therefore disregard Wittgenstein's opinions on mathematics.

As for Lakatos's ideas on mathematics (Lakatos 1976, 1978), they boil down to the following theses. Firstly, mathematical research is not essentially different from scientific research, for it too involves making conjectures and looking for counter-examples to them. Secondly, since one often starts with inexact concepts and one can make mistakes in proving theorems, one should adopt a fallibilist epistemology of mathematics. Thirdly, formalism does not depict faithfully the actual work of the mathematician, which involves nondeductive procedures. While all three theses are reasonable, neither of them is original, and they do not amount to a philosophy of mathematics. For one thing Lakatos had not clear ideas of his own about the nature of mathematical objects: he was more interested in the history than in the ontology or the semantics of mathematics. For another he exaggerated fallibilism to the point of skepticism, of denying that mathematics has any foundations, and of minimizing the role of logic – for all of which he earned the admiration of many a friend of sloppiness and inconsistency. So much for the marginal philosophies of mathematics. We turn now to the major classical schools.

Platonism is the philosophy associated with the logicist foundational strategy. (See e.g. Frege 1884, Gödel 1947, 1951, Wedberg 1955.) The best and shortest characterization of Platonism is due to Charles Hermite (of Hermite polynomials fame) in a letter to Stieltjes (of Stieltjes integral fame): "I believe that the numbers and functions of analysis are not the arbitrary product of our mind; I think that they exist outside ourselves with the same necessity as the things of objective reality, and that we meet, discover, or study them, just as the physicists, the chemists, and the zoologists" (art. "Realism", Lalande, Ed. 1938).

It is not hard to imagine the source of this philosophy and to explain its popularity. Trace a circle on sand: it will be imperfect compared with our idea of a circle, so you may have to retrace it or apologize for your clumsiness. Moreover your work will soon be wiped out by wind or water, whereas your mind will retain the idea of a perfect circle – an impersonal and universal idea. Ideas, then, and particularly the well defined ones of mathematics, seem to be perfect, endurable, and universal – unlike their material "embodiments", which are but imperfect, ephemeral, and circumstance-bound copies of the former. What can sound more self-evident, or better grounded in experience, than Plato's theory of Ideas (or Forms) and its application to mathematics?

No wonder that Platonism is the spontaneous philosophy of mathematicians: that most of them believe that they explore and eventually discover objects "existing independently of our definitions and constructions" (Gödel 1951 p. 137). Thus one mathematician asserts – without hinting at any possible evidence – that a mathematical structure "exists independently of any mind, and would exist even if there were no mind" (Goodman 1983 p. 65). And in a widely publicized interview, the Princeton professor William Thurston stated that "Theorems just kind of exist, you know, just like mountains do". In our view this is an intelligent mistake. It is a mistake because formal existence is radically different from material existence: recall Sect. 2.1. But it is intelligent because, as a matter of fact, the mathematician behaves in many regards *as if* constructs existed by themselves. He can do so because mathematical constructs, though human creations, do not bear the stamp of their creators: they are impersonal or intersubjective (though not objective).

In addition to accounting for the universality and apparent objectivity of mathematics, Platonism has been attributed virtues which either are not such or are shared by other philosophies. Its most obvious virtue is that, by proclaiming the logical autonomy of mathematics *vis-à-vis* all other fields of knowledge, Platonism allows for the greatest freedom of mathematical research. This stands out in contrast to the limitationist policies of intuitionism and empiricism. However, to ensure research freedom we only need to demarcate formal from factual science the way we did in Sect. 1. In this fashion research freedom is protected at a far lower price, without having to buy the weird thesis of the independent existence of ideas, which contradicts physiological psychology.

A second virtue Platonism is often credited with is simplicity. But simplicity is more often an indicator of superficiality than truth. Mathematics calls for an ontology, a semantics, an epistemology and a methodology of its own, distinct from those of factual science or technology. For example, proving a theorem is in every respect quite different from testing a hypothesis in the laboratory. A third virtue Platonism is attributed is that it accounts for multiple independent discovery or invention: this would prove that ideas exist by themselves and are up for grabs. However, there is a scientific explanation for such coincidences: similar brains, trained similarly, and sharing a fund of knowledge, are likely to come up with similar solutions to the same problems.

Besides being less meritorious than often assumed, Platonism has a number of fatal flaws. Firstly, it makes no room for non-standard models of abstract theories and for non-standard logics – or at least it does not help us choose among the bewildering multiplicity of logics, all of which would coexist peacefully in the Realm of Ideas. Secondly, it makes no room for partial truths, which are the staple of approximation theory and factual science. Thirdly, and most important of all, there is no shred of evidence for the fantasy that there are ideas in themselves, i.e. separate from brain activity: this is just a dogma. (Recall Vol. 4, Ch. 4, or Vol. 5, Ch. 1.)

Despite all of the above-mentioned defects, most mathematicians behave like Platonists even while paying lip service to some other philosophy Thus Bourbaki, officially a formalist, believes in the "reality of mathematics" (Dieudonné 1970 p. 145). This is because Plato had grasped correctly the impersonality and universality of mathematics, as well as the immateriality and unchangeability of its objects, and because he allowed mathematicians utter freedom as long as they proceed rationally, in particular consistently. However, our own philosophy explains the same features of mathematics and grants the same creative freedom but, unlike Platonism, it is consistent with science, which does not recognize disembodied ideas.

Nominalism is the philosophy behind formalism. (See e.g. Leśniewski 1927–31, Chwistek 1932, 1949, Zinov'ev 1973, Field 1980.) It is a form of vulgar materialism (*a*) for treating mathematical and physical objects in one breath (i.e. for refusing to distinguish conceptual from physical existence), (*b*) for asserting that there are general signs but no general, let alone abstract, ideas, and (*c*) for holding that a general sign does not designate a whole (such as a line) or a natural kind (such as a species) but denotes severally the individuals it points to. Since mathematics handles symbols (rather than concepts), and since symbols are conventional, mathematics turns out to be just as conventional as chess. Consequently the mathematical formulas are true by virtue of the rules of the game, i.e. by linguistic convention or stipulation – hence in a very Pickwickian sense of 'truth'. Mathematics would thus be a vast language – but one that does not talk about any entities.

Nominalism has attracted the formalists, who have worked primarily on the more abstract branches of mathematics, where axioms are the main, and sometimes the only, guide. However, not even Hilbert, the arch-formalist, was a full-blooded nominalist, for he admitted abstract objects and he acknowledged the need for (*a*) a notion of formal truth – inherent in his contributions to model theory, and (*b*) interpreted (*inhaltliche*) axiom systems side by side with uninterpreted ones (Hilbert and Bernays 1968 p. 2). Nowadays nominalists are to be found only in the Russian constructivist school (e.g. Markow 1975) and among philosophers.

The nominalist thesis about symbols as the objects of mathematics was refuted long ago by Frege (1895), who remarked that, precisely because symbols are chosen conventionally, they are inessential: what matters are the objects they stand for. (It is only *par abus de langage* that we write expressions such as 'Let $\mathbb{N}$ *be* the set of natural numbers' instead of 'Let '$\mathbb{N}$' *designate* the set of natural numbers'.) In sum, mathematical objects are not linguistic objects. (More on the symbol-concept distinction in Vol. 1, Ch. 1; on the numeral-number distinction in Kneale 1972.) As for the nominalist thesis about the purely "formal" or syntactic nature of mathematics, it was refuted by the emergence of the semantics of mathematics (Fraïssé 1982). According to the latter the meaning of a "formal system" (abstract or uninterpreted theory) is the set of its models (examples), i.e. the "structures" (systems) in which the axioms of the "formal system" (abstract theory) are satisfied – i.e. are formally (though not factually) true. However, the strongest objection to nominalism is perhaps that it demands an intolerable mutilation of mathematics. If it were taken seriously we would have to give up all the non-denumerable sets, such as $\mathbb{R}$ and its geometric counterpart, the line, for most of their members cannot be named: they are concepts, not names. I.e., unlike the set of all possible names, ℝ is non-denumerable.

Finally, the nominalist thesis that there are only individuals is sound with regard to material things, whereas the Platonic thesis of the reality of universals (such as "number" and "continuous") is inconsistent with factual science. But it is unreasonable to generalize that thesis to all objects – an unavoidable generalization given the tacit belief, common to both nominalism and Platonism, that there is but one type of existence. (Recall Sect. 2.1.) Modern mathematics rejects *both* the nominalist denial of universals and the Platonic thesis that universals (in particular properties) pre-exist individuals. In fact, the very definition of a class or species in terms of a predicate joins the two basic notions of universal and of individual: For any predicate F, Fa iff $a \in \{x|Fx\}$. In short, whereas factual science has vindicated the individualistic thesis with regard to the external world – though not the nominalistic prejudices against wholes and properties – mathematics has overcome the moninalism/realism dispute over conceptual objects ("abstract entities"). True, contemporary radical nominalists such as Nelson Goodman (1956) do not accept this solution and prefer to reject all versions of set theory. By so doing they remain alienated from modern mathematics.

As we intimated a moment ago, the nominalist thesis about symbols entails *conventionalism*. Now, conventionalism is utterly false with regard to factual science (Vol. 6, Ch. 12, Sect. 1.1), but it does have a grain of truth with reference to mathematics. In fact, it is true that (*a*) doing mathematics involves (though is not identical with) laying down and obeying certain definitions and rules, rather than conducting empirical tests or proceeding only inductively, and (*b*) mathematical truth is not established by empirical tests or by intuition but by playing the rules of deductive inference.

However, mathematical conventionalism has the following fatal flaws: (*a*) no mathematician ever adopts an axiom by arbitrary stipulation: axioms are adopted because of their fertility, and theorems are accepted when proved – and proved when interesting; (*b*) whereas linguistic conventions form a set devoid of a definite mathematical structure, mathematical statements tend to cluster into systems (theories); (*c*) no set of linguistic rules can help describe the world, whereas a mathematical theory, when enriched with suitable interpretation axioms, can; (*d*) unlike the rules of language, those of mathematics rest on laws (e.g. the factorization rule for numbers rests on the associative law); but not every law is a rule (e.g. an existence theorem does not tell us how to do anything: it just backs up our search for the object satisfying the theorem); (*e*) unlike chess and other games, "The best mathematics is *serious* as well as beautiful [...] The 'seriousness' of a mathematical theorem lies [...] in the *significance* [importance] of the mathematical ideas which it connects. [...] Thus a serious mathematical theorem, a theorem which connects significant ideas, is likely to lead to important advances in mathematics itself and even in other sciences" (Hardy 1967 p. 89). In conclusion, nominalism is not an adequate philosophy of mathematics: it does not yield a faithful description of mathematics, and it has no heuristic power.

Next comes mathematical *intuitionism* (Brouwer 1975, Weyl 1949, Heyting 1956b, Bishop 1967, 1975, Dummett 1977). Unlike Platonists, intuitionists hold that the objects of mathematics are creations of the human mind. (Some intuitionists have even sketched an interesting theory of the creative subject who abides by the intuitionistic strictures: see Martino, 1982.) This ontological and psychological thesis should be adopted by anyone who is not an objective idealist, and it fits snugly into our naturalistic ontology. But admitting this thesis does not commit oneself to accepting the epistemological thesis that every genuine concept embodies an intuition of some sort. Indeed, whether or not a statement is intuitive is a highly subjective matter – partly because of the very ambiguity of the term 'intuition' (Bunge 1962a). For most people untutored in mathematics the following ideas are counterintuitive: that $1 = 0.999\ldots$ (which results from adding 1/11 and 10/11, and taking the decimal expansions of these fractions); that there are infinitely many fractions between any two given fractions, however close; that the rationals are denumerable; that the numerosity (cardinality) of the rationals is as nothing by comparison with that of the reals; that there are non-associative objects, such as the Poisson brackets; that there are non-commutative operations, such as matrix multiplication; that there are curves, such as Peano's, that can fill a square. The thesis of the absolute character of intuitability is then empirically (psychologically) false. And the thesis that no unintuitive concept is valuable is dangerous because it hinders abstraction. It is not by chance that Grassmann's *Ausdehnungslehre* (1844), an early showpiece of abstract algebra and the origin of the vector calculus, was attacked by the Kantian philosophers of the time for being purely conceptual and unintuitive.

The methodological thesis of intuitionism, namely that the only legitimate way of introducing mathematical constructs is by explicit construction, is a horse of a different colour: in fact it is a respectable strategy of theory construction. In Sect. 5.1 we argued that it is an admissible foundational strategy provided it not be allowed to ban any alternative strategies. But the constructivist thesis has no philosophical justification, for there is no *ontological* difference between "real mathematical existence" (proved by construction) and "pure mathematical existence" (generated by axiomatic definition or proved by *reductio ad absurdum*). In fact both are cases of conceptual or formal, not physical, existence. In this regard an individual number – be it small like 9 or enormous like $99^{99^{99}}$ –, the set of real numbers, and even the power set of the latter, are on the same footing: *all of them are fictions* (or ideal objects in the Platonic terminology). True, there are psychological, epistemological and methodological differences between those various types of mathematical object. But some of these differences are subjective and are being erased. For example, "enormous" numbers occur not only in some mathematical proofs but also in statistical mechanics and in cosmology. (E.g. the Poincaré recurrence time for 1 cc of gas is of the order of $10^{10^{19}}$ years.) In any event, getting down to manageable numbers takes us nearer familiar ground but not nearer reality: all mathematical objects are physically unreal. And numbers, whether large or small, are just as meaningless as sets, infinite or finite, because they are neither predicates nor propositions.

Intuitionists make much of the word 'meaning' in a rather vague sense that is neither the one used in model theory nor the one occurring in factual science. For instance Bishop (1975), following Brouwer, indicts "classical" mathematics for being largely "meaningless", and he insists quite rightly that one should not discuss matters of truth before having settled questions of meaning. And Stolzenberg (1970 p. 318) asks us to believe that "the principle of mathematical induction is correct *simply by virtue of its meaning*" – presumably because it refers to natural numbers. But neither Bishop nor Stolzenberg tells us what he means by 'meaning', aside from vague remarks about computability. (Besides, there seems to be a contradiction in Bishop's remark because, according to the operationist semantics adopted by intuitionism, nothing makes sense apart from testing procedures. On this view the meaning of a formula is the way it is put to the test: truth is the source of meaning rather than presupposing the latter. But this thesis is as false in mathematics as it is in physics: see Bunge 1973a, Ch. 1, and Vol. 2, Ch. 7, Sect. 5.1. The content of a statement cannot be identical with the way it is validated. For example, the Pythagorean theorem is not the same as the 200 or so different proofs of it.) In conclusion, although mathematical intuitionism contains some valuable epistemological insights, it is not an adequate comprehensive philosophy of mathematics. Therefore those mathematicians who call themselves 'intuitionists' do well to dissociate themselves from the initial philosophical motivations of intuitionism. (See Bunge 1962a for further criticisms.)

Finally there is *mathematical empiricism*, which is shared by most dialectical materialists. (See e.g. Enriques 1913, 1940, Casanova 1947, Borel *apud* Fréchet Ed. 1967, Kálmar 1968, Davis 1974, Putnam 1975, Kuyk 1977, Lakatos 1978, Restivo 1983, Kitcher 1983, Harsanyi 1983.) Empiricism can be historical, philosophical, or both. The historical thesis of empiricism is that every construct originates, however remotely, in experience, so that there can be no such thing as pure mathematics. By leaving the adverb 'remotely' undefined, the thesis is irrefutable, for we can never be sure that there are no subtle threads linking even the most sophisticated abstract constructs to some sense experience. But taken literally the thesis is refuted by the great many cases where mathematical research was sparked off by problems of a purely conceptual kind: think of Diophantine equations, complex numbers, abstract algebra, non-Euclidean geometries, topology, or the theories of sets, real numbers, functional spaces, or categories. In all these cases the driving forces were curiosity and the wish to attain maximal generality or systemicity while preserving consistency. In short, historical empiricism is false.

The *philosophy* of mathematical empiricism boils to two theses. One is the semantical thesis that every mathematical object represent possible experiences or even "qualities of the world we live in" (Kuyk 1977 p. 137). The other is the methodological thesis that mathematical research proceeds much in the same way as empirical research, namely by trial and error, analogy, and induction, and that the ultimate criterion of mathematical truth is experience. A classical statement of the semantical thesis is found in Mill (1843 Bk. iii, Ch. xxiv, Sect. 5): "Each of the numbers two, three, four, & c., denotes physical phenomena, and connotes a physical property of those phenomena. Two, for instance, denotes all pairs of things, and twelve all dozens of things, connoting what makes them pairs or dozens; and that which makes them so is something physical". A more recent example is the attempt to reconstruct set theory in terms of concrete operations on perceptible things, such as collecting, ordering, and matching (Kitcher 1983 pp. 126 ff.)

To be sure there are a few examples like Mill's, which works for small positive integers, or like Kitcher's, which works for small collections of perceptible objects. But they are limited to the *early* stages of *some* branches of mathematics. There are many other cases where there is no reference at all to physical objects or to empirical procedures. Think of irrational numbers (which cannot be the outcome of measurements), of continuous functions without derivatives (hence incapable of representing possible movements), of infinite (in particular divergent) series, or of abstract spaces. Besides, if every mathematical theory were just a summary and codification of experience, there should be but one arithmetic, one geometry, one set theory, one analysis, one logic, and so on. In fact there are many mathematics and a single world.

Even granting that *some* mathematical constructs have originated in sense experience or in manual practice, it does not follow that the proper way of investigating them is empirical. To hold this view is to incur the genetic fallacy. Geometry is not an empirical science just because it originated in land surveying, anymore than number theory employs empirical methods just because it started as counting fish or shells. If mathematics were an empirical science, theorem proving should be empirical not conceptual: for example, we should claim that the physical addition of things proves the laws for the addition of numbers – but then we should overlook counterexamples such as the juxtaposition of two liquid droplets, which may result in a single droplet, and the high energy collision of two elementary particles, which may result in a dozen particles. So, it is false that "the criterion of truth in mathematics just as much as in physics is success of our ideas in practice" (Putnam 1975 p. 61).

(Platonists, and perhaps moderate empiricists, would deny that we can "read mathematics off science", but they would argue that the interlocking of mathematics and factual science suggests that mathematics is about the world after all. Thus the eminent MacLane (1981 p. 471): "Mathematics consists in the discovery of successive stages in the formal structures underlying the world and human activities in that world, with emphasis on those structures of broad applicability and those reflecting deeper aspects of the world". In our view, the facts that mathematics is the heart of modern theoretical factual science, and that the latter has often stimulated the former, prove not that mathematics is just as *empirical* – in origin, essence, and method – as factual science, but that the latter is *rational* as well as empirical. In any event, either thesis has got to be justified by a semantical analysis of the reference of mathematical concepts such as "$\in$" and "d/dx". We have done this job in Vols. 1 and 2, which show that $\in$ refers to sets in general and d/dx to functions in general. On the other hand Platonists and empiricists still owe us viable semantic theories capable of disclosing the referents of any predicates or propositions, and of substantiating their claim that mathematics studies the structure of the world. And neither can possibly account for the recent finding that there are at least two different systems of real numbers, so that reality contains neither of them.)

The doctrine that mathematics mirrors human experience, or even the world, is open to the following additional objections: (*a*) no mathematical proof represents anything in reality, if only because the deducibility relation has no material counterpart; (*b*) by itself no mathematical formula represents anything in reality: so much so, that one and the same formula, such as a differential equation, can be assigned a number of alternative factual interpretations; (*c*) take a process such as a macrochemical reaction, on the one hand, and the equation in chemical kinetics representing that reaction, on the other: the equation contains a parameter (the rate constant) that must be determined empirically, and every term in the equation is assigned a factual reference by an extramathematical assumption ("correspondence rule"); (*d*) whereas physical (or chemical, biological, or social) links are factual (think of pushing or burning), mathematical relations are not: they have no causal efficacy.

Yet in recent times it has become somewhat fashionable to claim that mathematics, even logic, can be "read off science". In particular it has been held that quantum mechanics has refuted standard logic in the microphysical domain. (Recall Sect. 3.2.) Most authors rest this claim on an examination of a couple of experiments such as the double slit diffraction experiment. But, far from carrying out the calculation within the framework of an axiomatic formulation of quantum mechanics – in order to be able to ascertain what the underlying logic is – they take the theory as usually formulated and work from then on. Had they proceeded rigorously, i.e. by first axiomatizing the theory, they would have realized that it makes exclusive use of theories in classical mathematics, which in turn presuppose classical logic. (See e.g. Bunge 1967c.) Hence any seeming violation of ordinary logic in microphysics must be a mistake in the application or the interpretation of the mathematical formalism of quantum theory. (Recall Sect. 3.2.) In conclusion empiricism is a false philosophy of mathematics – except for the platitude that mathematical work is a type of human experience. (Further criticisms in Goodstein 1969, Goodman 1979, Bell and Hallett 1982.)

This concludes our critical survey of the four classical philosophies of mathematics, which will be found summarized in Table 1.1.

Every one of the classical philosophies of mathematics has some good points, but none of them covers adequately all aspects of mathematical research: posing and reformulating problems, using theories or guessing

TABLE 1.1. Typical tenets of some philosophies of mathematics.

Philosophy	Math. objects	Mode of introduction	Meaning	Truth	Math. knowledge	Math. activity
Platonism	Self-existing ideal and eternal	Discovery	Non-contradiction	Formal	A priori and conceptual	Deductive
Nominalism	Symbols	Convention	Nil	Convention	Nil	Formal manipulation of symbols
Intuitionism	Mental constructions	Invention	Reducibility to positive integers	Reducibility to numerical computation	A priori and intuitive	Intuitive and rational
Empiricism	Mental	Discovery	Reference to experience	Empirical	Empirical	Trial and error, rational and empirical
Conceptualist and fictionist materialism	Fictions (classes of brain processes)	Invention and discovery	Conceptual reference and contextual sense	Formal	A priori and conceptual	Abstraction, generalization, formal manipulation, trial and error, analogy, induction and deduction

hypotheses to solve them, proving theorems, inventing axioms, definitions and algorithms, computing, comparing constructs, making mathematical considerations, etc. – all the while using intuition, analogy, induction, and deduction. Our own philosophy of mathematics, sketched in the preceding pages, was intended to retain those valid points and to cover the main aspects of mathematical research as well as of its finished products. In fact

(i) it accounts for the purely conceptual nature of (finished) mathematical objects and methods, while making room for the empirical or intuitive origin of some of them;

(ii) it accounts for the universality and impersonality of mathematical constructs while stressing that they are brain-children likely to be created only in fairly advanced societies;

(iii) it accounts for the differences between formal and factual propositions, as well as between mathematical proof and empirical validation;

(iv) it accounts for the difference between an abstract theory and its models as well as between logical models and the models in science and technology;

(v) it accounts for invention (of new constructs), discovery (of logical relations), and routine work (in applying rules or algorithms);

(vi) it accounts for the applications of mathematics without rendering it the handmaiden of science and technology: service but not servitude;

(vii) it leaves maximal freedom compatible with consistency and systemicity (relatedness to other ideas);

(viii) it respects the logical stratification of mathematics (elementary logic, category theory, set theory, number theory, abstract algebra, topology, analysis, etc.);

(ix) it requires neither mythical objects, such as self-existing Platonic ideas, nor non-rational faculties, such as intuition (except as a heuristic aid);

(x) unlike all other philosophies of mathematics, it is part and parcel of a comprehensive philosophical system consistent with science and technology.

6. CONCLUDING REMARKS

Like any other human activity, mathematical research and its product can be looked at in a number of alternative ways. One is the logical viewpoint, which regards mathematics as a finished product and, more particularly, as a system of theories. Another is to focus on mathematical activity – its motivations, modes of reasoning, etc. – as a particular kind of mental (or brain) activity; this is the psychological viewpoint. A third is to look at mathematical research as a type of social activity, and at its product as a special kind of cultural artifact: this is the sociological viewpoint. A fourth is to study mathematics as a historical process of discovery, invention, and diffusion throughout a given society: this is the historical viewpoint. A fifth is to study mathematics as a tool for factual science, technology, and the humanities: this is the instrumental viewpoint. A sixth is to look upon mathematical research as a particular mode of cognition, and its product as a special kind of knowledge: this is the epistemological approach.

These different ways of looking at mathematics are mutually compatible, nay complementary. Therefore it would be wrong to espouse one of them and exclude all the others, for mathematics is simultaneously everything those alternatives say it is. However, the epistemological viewpoint is methodologically prior to the others because it is the one that tells mathematics from other disciplines. (Imagine trying to trace the history or the social ties of a mathematical idea without first making sure that it is indeed *mathematical* rather than scientific, technological, or philosophical.) Still, the philosopher cannot do a good job of characterizing mathematics unless he takes the other points of view into account. A couple of examples will show the need for such extraphilosophical information.

Example 1 Consider the formulas "$2 + 3 = 5$" and "$5 = 3 + 2$". They are mathematically identical – except of course to the nominalist, since they are typographically different. But they are psychologically (or pragmatically) different: the first may be interpreted as the instruction to "take" (think of or handle) first the number (or the numeral) 2, then add it 3; on the other hand the second formula may be construed as an instruction to decompose 5 into 3 and 2 (in this order, not the reverse). The possibility of such an extramathematical interpretation helps explain the survival of the empiricist (in particular pragmatist or operationist) philosophy of mathematics. *Example 2* At first sight mathematics is the lonely profession in the sense that it is rarely the work of teams, it is unpopular, and it is hard to popularize. However, mathematics is just as dependent upon tradition and communication as any other research field. It is not only that the information channels must be kept open in order for mathematics to flourish, but that some of its key ideas seem to have been generated in social intercourse. Thus the idea of a proof may have been necessitated by the need to persuade somebody (Hammond 1978 p. 30). And the idea of consistency may have originated in the need to attain consensus in a political democracy. (Politics, not the economy, explain why logic was born in Greece rather than in any of the despotic slave states.) This is no mere matter of historical erudition: philosophers should know more about the psychological and social mechanisms of validation of truth claims in order to understand why the rules of inference and the standards of rigor are anything but absolute.

Given the nature of this book we have adopted an epistemological approach to mathematics. However, from the beginning we have stressed the anti-Platonist thesis that mathematics does not exist except in the brains of some people living in societies that encourage or at least tolerate free speculation and the free discussion of ideas. How can this materialist view jibe with the fictionist component of our philosophy of mathematics, according to which when creating or utilizing a mathematical construct we *feign* that it leads an impersonal and suprasocial existence? There is no contradiction here, for we hold that it is we, living beings immersed in a concrete society with a rich tradition, who construct such fictions. (For a critical examination of our brand of fictionism see Torretti 1982.)

Still, is not such fictionism regarding mathematics inconsistent with the scientific realism espoused in Vols. 5 and 6? No, because our fictionism, unlike that of Nietzsche or Vaihinger, is limited to fictions such as mathematical systems and mythologies: it does not carry over to factual science and technology. In other words, our epistemology is realistic concerning the study of the real world, and fictionistic concerning fictions. It is thus dualistic in that it preserves Leibniz's distinction between *propositions de raison* and *propositions de fait* (Sect. 1.1). Hence it is more complex than its rivals, most of which are monistic in assuming that there is a single kind of existence, hence of proposition. Yet, epistemological dualism is not self-contradictory. It is consistent for it does not predicate fictitiousness of factual scientific constructs (e.g. the equations of chemistry) or material reality of mathematical fictions. (That there is often a correspondence between them, as between atoms and the equations representing them, is another story.) Moreover such epistemological dualism does not carry over to our ontology: this one is thoroughly monistic because it does not postulate that constructs are part of the furniture of the world. What are found among the furnishings of the real world are brains capable of creating constructs.

There is much more to the philosophy of mathematics, but we must keep going. Henceforth we shall only encounter mathematical objects as components of our ideas about reality.

CHAPTER 2

PHYSICAL SCIENCE FROM PHYSICS TO EARTH SCIENCE

We start now our study of some topical problems in the philosophy of factual or empirical science. As we saw in Ch. 1, factual science is semantically and methodologically quite different from mathematics. Unlike mathematics, factual science seeks truths of fact – i.e. propositions representing reality – and it employs not only mathematical reasoning but also empirical procedures. However, factual science and mathematics share a number of traits, and this is why both are rightfully called 'sciences': both employ the general scientific method, both attempt to maximize exactness and systemicity, both justify their assertions, and both change relentlessly as a result of research rather than in obedience to dogma or social pressure. (Recall the general definition of a scientific research field in Ch. 1, Sect. 1)

We begin with the philosophy of physics for the following rather obvious reasons. Firstly, physics is the most basic and universal of all factual sciences, because (*a*) every real thing is either a physical entity or is composed of physical entities, and (*b*) the laws of physics are not place or time bound. Secondly, physics (including astronomy) was the first discipline to attain a scientific status and, as a consequence, it showed the way of science to all the other research fields. Thirdly, the philosophy of physics (including astronomy) is the oldest, most firmly established and most flourishing of all the philosophies of science. (Ptolemy was one of its founding fathers, and Galilei its first modern practitioner.) For this reason the philosophy of physics showed the way to the other regional philosophies of science. This has been both a blessing – because of the high standards it required – and a misfortune – because we have tended to overlook the specific differences between the physical sciences and the others. (Biologists often complain that philosophers of science treat their science as if it were a mere extension or application of physics. See Ch. 3, Sect. 1.1.)

Although the philosophy of physics is the most advanced branch of the philosophy of science, it is still underdeveloped in a number of respects. For one thing much of it is based on secondary literature. Secondly, philosophers of physics are often overspecialized – e.g. in the philosophy of space and time, or of quantum mechanics. Consequently they tend to take problems out of their general scientific setting, which prevents them from controlling the proposed solutions by checking their compatibility with the bulk of physics. Thirdly, philosophers of physics seldom attack problems in the light of an explicit semantics, epistemology, methodology, and ontology. Consequently they often come to conclusions that are at variance with the marrow of physics – witness the views that the world is a system of mathematical equations, that there are no laws but only conventions, that there is a general and a priori theory of measurement, that change is illusory, or that physics does not study autonomously existing things. Fourthly, although many workers in the philosophy of physics master some of the mathematics of physics, they seldom make use of the mathematics of philosophy. As a consequence their assertions about exact formulas are often inexact for including unanalyzed key philosophical or methodological concepts, such as those of matter, event, causation, physical quantity, factual meaning, factual truth, and measurement.

There is, then, much to be done in the philosophy of physics. In this chapter we shall grapple with a handful of open problems, and shall do so by availing ourselves of some of the tools wrought in previous volumes of this Treatise. On occasion we shall also make use of results, in physics and its philosophy, obtained by the author or his coworkers. The first Section will be a *pot pourri* devoted to elucidating half a dozen key concepts of physical science. These concepts are so familiar that they are seldom analyzed – a sure sign of their scientific importance and philosophical complexity.

1. PRELIMINARIES

1.1. *Physical Quantity, Convention, Measurement*

Physical science studies physical things such as bodies and fields. (This platitude is of course denied by subjectivists, operationists and conventionalists.) Physical things can be simple – i.e. without separable components – though possibly extended; electrons and photons seem to be both simple and extended. Or they can be complex, i.e. systems proper such as atoms and clouds. All things other than physical things are systems, and they exist only in special environments. For example, chemical systems are confined within rather narrow temperature and pressure bounds, biosystems exist only in a narrow range of habitats, and sociosystems under even more special circumstances. (Recall Vol. 4.)

The properties of a physical thing are appropriately called 'physical properties'. If a thing has properties other than physical properties then it is not a physical thing and therefore it lies beyond the reach of physical science. (The latter may on the other hand study the physical components of a supraphysical thing.) This truism is at variance with the subjectivistic interpretation of certain physical theories, particularly quantum theory. According to that interpretation there are no physical things independent of human thought and action. Consequently physics would be indistinguishable from psychology or even epistemology: it would study human knowledge rather than nature. This view, which stems from Berkeley and Mach, will be examined in Sect. 4.2. Suffice it to state here that it is at variance with the practice of physical research. Indeed the theoretical physicist attempts to model physical things in terms of purely physical properties, and the experimental physicist racks his brains to keep himself at arm's length from the things he studies in order to minimize the disturbances he may cause.

All physical things, however simple, have numerous properties; the one property they all share is presumably energy. All the properties of a physical thing-in-its-environment are lawfully interrelated. The central goal of theoretical physics is to conceptualize such properties and the constant relations among them, i.e. the laws of physical objects. And the central aim of experimental physics is to measure physical properties, explore their inter-relations, and test plausible conjectures about the latter.

Every property is conceptualizable (representable by) at least one physical "quantity", whether constant or variable. In turn every such "quantity" is either a function (e.g. the temperature function) or an element of an algebra – e.g. the spin operator. (We distinguish a function from its values, and an element of an algebra from its spectrum.) For example, the value of the temperature of a macrophysical system in a given state (and reckoned on the absolute temperature scale) is some non-negative real number. Ordinarily, the values of physical properties are n-tuples of real numbers. But the converse is false. Thus spacetime is currently represented by a differentiable manifold, every point of which is identifiable (relative to a given coordinate system) by a quadruple of real numbers. But these numbers are not the values of any functions. This is not surprising, because spacetime is not a physical property of any particular physical objects: it is the public property of all. (On the other hand the spatial position and age of a physical object can be construed as the values of certain functions.)

It will pay to be more precise and examine a couple of examples. Our first example will be one of the simplest physical "quantities", namely electric charge. This is an intrinsic property of some bodies, i.e. one that does not depend upon any other physical object; in particular it is invariant with respect to changes in reference frames. Electric charge can be conceptualized as a function from the collection of all possible bodies to real numbers. (Note that (*a*) we employed the word 'collection', not 'set', for we are dealing with a variable collection not a fixed set; (*b*) we refer to all possible bodies, not just those that happen to exist right now relative to our time frame; (*c*) our collection of bodies includes not only those that happen to be electrically charged (the current extension of the concept of charge) but all possible bodies, for even the ones that are now electrically neutral may acquire an electric charge.) However, we can reckon or measure values of electric charges in alternative charge units, e.g. the e.s.u. and the coulomb. Therefore we must write "$Q(b, u) = r$" for "The value of the charge of body b, in unit u, equals the real number r". The general concept of electric charge is then

$$Q: B \times U_c \rightarrow \mathbb{R},$$

where B is the collection of possible bodies, U_c that of conceivable charge units, and $\mathbb{R}$ the system of real numbers.

(Note that the electric charge on a body is a fact, even though the unit in which it is calculated or measured is conventional. Also, it is a fact that positrons and electrons have the same charges with opposite signs, though calling the former 'positive' and the latter 'negative' is mere convention. All this is trivial. What is not trivial is the matter of the transformation properties of the electric charge, in particular its behavior under orientation reversal, i.e. $x \rightarrow -x$, and time reversal, i.e. $t \rightarrow -t$. The usual assumption is that electric charge is invariant under these operations, i.e. that it is an absolute scalar. However, it is occasionally assumed that Q is reversed with time reversal, or that Q is reversed with space inversion. The problem is still open and it cannot be solved by studying it in a single branch of physics, as the charge concept occurs almost everywhere. See Post 1977.)

A slightly more complex example of a physical quantity is that of position. We can write "$P(b, f, t, u) = \langle x, y, z \rangle$" for "The position of point-like object b, relative to reference frame f, at time t, and in unit u, is the ordered triple of real numbers $\langle x, y, z \rangle$". In general, the (classical) concept of position coordinate is a function of the form

$$P: B \times F \times T \times U_d \rightarrow \mathbb{R}^3,$$

where now B is the collection of point-like objects – or of points of an object, such as a wave-front –, F the collection of reference frames, T that of instants, U_d that of distance units, and $\mathbb{R}^3$ is the collection of ordered triples of real numbers. In the quantum theories there are several position operators. (See Kálnay, 1971.)

The following points of methodological or philosophical interest are worth mentioning. First, physicists distinguish physical properties from their mathematical representatives. This distinction is usually tacit, but it is explicit in the quantum theories, where one talks about the operator $\hat{A}$ representing the dynamical variable (or "observable") A. Obviously, such a distinction would make no sense at all if physics were wedded to an idealistic epistemology. (More on the property-attribute distinction in Vol. 3, Ch. 2, Sect. 1.1.)

Second, it is common for a given physical property to be represented by different functions or operators – or by none at all – in different theories. For example, relativistic mass, unlike classical mass, is frame-dependent; and in relativistic quantum mechanics the relative mass is represented by an operator. In other words, the "quantity"-property relation is many-one. (Equivalently: let $\mathbb{P}$ be a set of physical properties, and V the totality of "quantities" representing them. Then the representation function is of the form $\rho: 2^V \rightarrow \mathbb{P}$, where 2^V is the power set of V.) This fact shows the need for distinguishing between objective properties and the attributes or predicates representing them. The idealist and the naive realist do not need this distinction; by the same token they cannot account for its occurrence in physics.

Third, there is no mathematical difference between a physical "quantity" and a function (or operator or an element of an algebra). The difference is made in an extra assumption of a semantic nature. Thus in the case of the electric charge function Q treated above we assert that $Q(b, u)$ *represents* the electric charge, in unit u, of body b. In our terminology this is a *semantic assumption*, known as a "correspondence rule" in the literature. The difference in terminology corresponds to a philosophical difference: (*a*) whereas according to positivism a "correspondence rule" correlates symbols with empirical operations, in our view a semantic assumption correlates concepts with objective properties that may or may not be measured with different techniques; (*b*) whereas on positivistic metascience these correlations are rules or instructions for interpreting symbols (so as to endow them with an empirical content), on our philosophy they are hypotheses and, as such, they are empirically testable (though not in isolation). For example, the supreme "correspondence" rule of the Copenhagen interpretation of quantum mechanics is that every formula of this theory is about some experimental situation, none about anything in itself. (More in Sect. 4.2.) This empiricist principle is refuted every time a physicist makes a calculation referring to an isolated thing, or a measurement on an object lying beyond the reach of his laboratory.

Fourth, the mathematical representation of a physical property may contain conventional elements, such as units (sec, volt, degree, etc.). Only dimensionless magnitudes, such as relative increments and the probabilities of states, fail to contain units. (For a mathematical theory of dimensions and units see Bunge, 1971.) Units are not the only conventional components of physical theories. Another conspicuous component that can be chosen conventionally is the kind of coordinate system – e.g. rectangular, spherical, or cylindrical. (A coordinate system is not to be mistaken for a reference frame: the former is a concept, the latter a physical thing, such as a semi-rigid body with a clock attached to it. Any given reference frame can be represented by a number of alternative coordinate systems.) It might be thought that the occurrence of such conventional components refutes realism and confirms conventionalism. This is like saying that, since all painting includes conventional components, portraits have no real subjects; or that, since the GNP of a nation can be reckoned in dollars, yen, or rubles, there is no such thing as GNP. Unavoidably, every conceptual representation of a real thing is made up of concepts, and some concepts may represent things more or less faithfully, but they are not identical with the things represented.

Our analysis of a physical magnitude serves to identify its conventional component(s). And that analysis, coupled to our theory of reference (Vol. 1, Ch. 2), helps us exhibit the real referents of any physical magnitude. Thus, the real (or factual) reference class of the concept Q of electric charge, studied a while ago, is the collection B of possible bodies, which is properly included in the total reference class $B \cup U_c$ of Q. Similarly the real referents of the concept P of position coordinate are physical objects and, in particular, reference frames. (The total reference class of P is $B \cup F \cup T \cup U_d$. Since time does not exist by itself, and the distance units are conventional, we discard T and U_d when forming the factual reference class of P.)

Fifth, all things have a number of properties, presumably a finite number of them. (However, the values of these properties are different in different states – by the very definition of "state" – and relative to alternative reference frames. For example, the mass of a body has as many possible values as the values of its velocity relative to all the possible inequivalent reference frames.) But not all properties are equally important or of equal interest for a given purpose. Hence the problem of choice. This problem is solved differently by people with different interests. Thus the technician and the man in the street prefer things with a small number of easily measurable or controllable properties – such as the three variable system in Figure 2.1. A state of this system is a consistent triple $\langle a, b, c \rangle$ of angular positions of the pointers showing the values of the properties A, B, and C. On the other hand the experienced scientist knows that things are always more complicated than they look at first sight, and that the most important properties are often the most difficult to measure or control. So, he may investigate any "well known" thing and, with talent, industry and luck, he is likely to discover further properties of it. If he is a theorist he will not care much for ease of measurement or control – which are on the other hand prime considerations of the technologist – because he knows that they depend upon the state of the art as well as upon the thing itself. He knows that certain properties, such as spin, stress, and space curvature, are hard to measure as well as important.

An important or basic property of a thing is one upon which other – derivative – properties depend: recall Vol. 3, Ch. 2. Important properties, or rather the magnitudes representing them, occur or should occur in law statements such as equations of motion and field equations. And a property is the more important, basic, or fundamental, the more fundamental the law statement(s), or their deductive consequences, in which its representative occur(s). In turn, the degree of fundamentalness of a law statement in a theory is given by its rank in the theory. Thus a variational principle is more basic than the equations of motion, or the field equations, it entails; and the latter are more basic than any of their solutions. As with properties so with law statements: the most general among them are the more important or

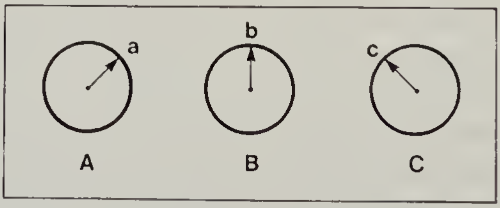


Fig. 2.1. Control panel of a three variable thing. The thing happens to be in state $\langle a, b, c \rangle$. The values a, b, and c form a consistent triple in that they are related by the law(s) of the thing.

basic. For example, the general equations of motion in continuum mechanics are more important than the constitutive equations, which inter-relate some of the properties of bodies of special kinds. However, no realistic special problems can be solved without using some such special laws, and the technologist can often get by with the sole help of the latter. Thus the petroleum engineer may set more store on the constitutive equations that characterize oil than on the general equations of hydrodynamics; likewise the electrical engineer may find more use for the special laws of circuits, such as Ohm's and Kirchhoff's, than for Maxwell's field equations. This is just one example of the difference between what is scientifically important and what technologically important: between knowing for its own sake, and knowing for the sake of doing. See Vol. 6, Ch. 14, Sect. 2.2.)

Sixth, exhibiting the general mathematical form of a magnitude, e.g. as a real valued function of two variables (i.e. $F: A \times B \rightarrow \mathbb{R}$), tells us very little about the magnitude. We also need to know the so-called rule of assignment of values to every point in the domain of the function (i.e. what exactly is the value of the F above at the point $\langle a, b \rangle$ in $A \times B$). In physics these "rules" are the law statements linking the given magnitude with other magnitudes. For example, Ohm's law, a constitutive equation, states that the possible values of the electric current in a d.c. metallic circuit are proportional to the impressed electromotive force. In general, the (nomologically) possible values of a magnitude are determined by the entire set of hypotheses (e.g. equations) that define (implicitly or explicitly) the magnitude.

Seventh, the law statements determine only possible values of magnitudes. The actual values depend on the actual special circumstances and they must be found out by measurement. As we saw in Vol. 6, Ch. 11, Sect. 3.1, measurement can be direct or indirect, i.e. with the help of well confirmed hypotheses, among which indicator hypotheses play a crucial role. For example, the length and period of oscillation of a simple pendulum can be measured directly (though not very accurately). And these values can be plugged into a well-known formula to yield the local acceleration of gravity.

Let us have a closer look at physical measurement, for it poses some philosophical problems that have hardly been explored. Whereas theoretical physicists deal most of the time with physical objects in themselves, such as protons and galaxies, regardless of the ways they can be observed or experimented on, experimental physicists handle physical objects "for us" (the knowing subjects). Indeed, they design, assemble (or construct), and operate measurement systems in which the physical objects of interest are included as components. We recall from Vol. 4, Ch. 1, that a system may be characterized by its composition, environment, and structure. In particular, a *measurement system* may be described as follows:

(i) *composition*: a *mensurandum* or object of measurement (e.g. an atom or a light beam), an *apparatus* (e.g. a seismograph), and an *experimenter* (or his automated proxy);

(ii) *environment*: an artificial milieu the main properties of which are carefully monitored and controlled, and that is maximally shielded from perturbations coming from the experimenter as well as from the part of the world that is not being studied;

(iii) *structure*: the collection of links among the system components, and among these and environmental items. In a well-designed measurement system, the salient interaction is that between mensurandum and apparatus: the influences of the experimenter and of the environment are negligible or can be corrected for. More on this anon.

In turn, the apparatus can be analyzed into the *measuring* instrument(s) – e.g. a pendulum – and the *recording* device(s) – e.g. a pen attached to a pendulum and writing on a strip of moving graph paper. The measuring instrument must be sensitive enough to detect small changes in the state of the mensurandum. And, when triggered by the latter, the instrument should be capable of causing an *irreversible amplification* in the recording device – e.g. an ink trace, a spot on a photographic plate, or an electric discharge.

It goes without saying that both the detection and the recording processes are assumed to be purely physical. In particular, psychokinetic effects are excluded by hypothesis for, if admitted, the measurement would be rigged and therefore worthless. In general, the role of the experimenter is limited to the design and operation (personally or by proxy) of the measurement system. Once the measurement system has been set up, the experimenter is supposed to keep at arm's length from it; moreover he must be able to be replaced by any other competent individual. Otherwise the measurement results would not be replicable and consequently they would be suspect.

Since the experimenter is not a permanent feature of a measurement system, any more than the mensurandum, and since the mensurandum-apparatus interaction is assumed to be strictly physical, it is mistaken to claim (as does Wigner 1963) that consciousness plays an essential role in the measurement process. If consciousness did play an important role in measurement, the latter would not be measurement. As Wheeler (1978) put it, "not consciousness but the distinction between the probe and the probed" is central to scientific observation. This distinction is denied by the standard quantum mechanical theory of measurement (von Neumann 1932, London and Bauer 1939), according to which the border between the knowing subject and the physical object is arbitrary. More on this point in Sect. 6.1.

Although every competent experimentalist takes pains to avoid interfering personally with the measurement process, he may deliberately design or use measuring or recording devices that produce strong perturbations on his mensurandum, to the point of altering it qualitatively. For example, to localize a photon he may resort to a photographic plate that will absorb it. And when using a Stern-Gerlach apparatus to measure electron spins, he actually causes the polarization of the incoming electrons, which are normally unpolarized (i.e. in a superposition of spin eigenstates). In fact the magnetic field of the Stern-Gerlach apparatus (but not the observer by himself) projects the spins onto the field axis, some in the same direction, others in the opposite one. (More in Sect. 6.1.)

Needless to say, any disturbances caused by the apparatus on the mensurandum are strictly physical, hence describable with the help of physical theories: they are not effects of mind on matter. Nor do they prove that scientists study only what they make, let alone that they conjure up all the events in the world. Whatever changes are caused by a measurement apparatus on a mensurandum are not only strictly physical but also strictly lawful: there is nothing arbitrary about them. (The projection of the state function is an apparent exception. We shall see in Sect. 6.1 that it is a consequence of lawful interactions.)

In other words, there are limits to what the experimentalist can do. For example, he cannot transform an electron into a neutron, or bring a photon to rest without destroying it, or change an invariant property (such as electric charge or entropy) by merely adopting a different reference frame. The Copenhagen interpretation of quantum mechanics denies both the objectivity and the lawfulness (hence the inherent limitations) of microphysical events: it asserts dogmatically that all of the quantum events are freely "conjured up" by the experimentalist (e.g. Bohr 1934, Rosenfeld 1953).

The design, performance and interpretation of precision physical measurements calls for *measurement theories*. There are two families of such theories as regards generality and testability. The first, invented by mathematical physicists and widely discussed by philosophers, is composed of extremely general theories involving the principles of quantum mechanics. They have no classical correlates and they attempt to account for the measurement of arbitrary physical magnitudes with the help of nondescript apparata and arbitrary techniques. Being so unrealistic, these theories make no specific predictions and consequently they remain untestable – which is a very convenient feature because they can be multiplied *ad libitum* and discussed to no end.

The second family of measurement theories, produced by theoretical physicists or experimentalists turned theorists, is composed of specific theories. They are usually in the *Review of Scientific Instruments* rather than in textbooks. (Since they are all based on general theories, such as mechanics, wave optics, or thermodynamics, they are bound models in the nomenclature of Vol. 5, Ch. 9, Sect. 1.2.) These theories account for the way particular properties (e.g. acceleration, viscosity, electric field intensity) are measured with the help of particular pieces of apparatus. Such specific theories are classical, quantum-mechanical, or semiclassical. *Examples*: the theory of Millikan's oil drop method for measuring the charge/mass ratio; the theory of the way the Geiger counter works; the theory of the mechanism whereby a photographic plate records the passage of a charged particle. Unlike the general measurement theories, the specific ones are testable and, moreover, they are used to design and operate measurement systems, as well as to improve them. But, since we have started to talk about theories, we are already trespassing on the next section.

1.2. *Theory, Metatheory, Protophysics*

Contemporary physics is essentially composed of theoretical physics and its assistant, mathematical physics, together with experimental physics and its own assistant, instrumentation. Physicists usually specialize in either of these fields. (Enrico Fermi was probably the last great physicist who felt at home in all four subfields.) The specialist in instrumentation and the mathematical physicist ignore each other, but the theoretical physicist is supposed to hear about experimental results once in a while; and the experimental physicist must have an intuitive grasp of a number of theories, for he uses some of their results to design and interpret his experiments. Because philosophers have a theoretical cast of mind, they have generally overlooked the rich philosophical problematics of experimental physics and have concentrated instead on a few theories, mainly classical particle mechanics, the two relativities, and quantum mechanics. (Continuum mechanics, statistical mechanics, classical electromagnetic theory, quantum electrodynamics, theoretical solid state physics, and many other theories are still waiting to be explored philosophically.)

The theories and theoretical principles of physics are not only numerous but they are also of different kinds. (All theories contain principles, or postulates, but not all principles are components of theories. More on this below.) In fact we may distinguish the following levels of physical theorizing:

(i) *Protophysical* principles and theories (e.g. physical geometries)
(ii) *General theories*
  (*a*) fundamental (e.g. classical mechanics)
  (*b*) phenomenological (e.g. classical thermodynamics)
(iii) *Special theories*
  (*a*) about natural things
  (*b*) about artifacts (e.g. measurement systems)
(iv) *Metatheoretical principles*

We call *protophysics* a motley collection of more or less tacit principles, as well as of explicit theories, that underlie theoretical physics (Bunge 1967c). Because protophysics belongs in the foundations of physics and is largely tacit, it behooves the philosopher to ferret it out, analyze it, and systematize it. We have done some of this job in Vol. 3, which was devoted to things in general, i.e. regarded as physical objects. But there we did not touch on any of the rather special protophysical principles presupposed by the very methodics of physics. (See Bunge 1967c for some of them.) Let us now have a look at one such principle, which has played a central role in the recent philosophy of space and time: the isometry hypothesis. But before we do so we must recall what a measurement standard is.

All physical measurements involve the use of *standards*, e.g. of time or of mass, which embody units of magnitudes. Standards are either things (such as the classical mass standard) or processes (such as a time standard). They are natural, such as a spectral line, or artificial, such as a battery. All standards are easily observable or replicable, and they are assumed to be constant under specified circumstances. Thus an atomic length (or time) standard is assumed not to change in the course of time provided the atoms that compose it are kept at constant temperature and pressure. Being factual items, standards are describable but not definable. On the other hand units (e.g. of time) are conventional; moreover, the derived units are definable in terms of the basic ones with the help of law statements. (For example the newton, a force unit, is definable via Newton's second law of motion in terms of certain units of mass, length, and time.)

According to Mach, Poincaré and their followers – in particular Reichenbach and Grünbaum – standards are thoroughly conventional. In particular, the length constancy (isometry) of a length standard would be convention not fact. This opinion, crucial to the conventionalist philosophy of space and time, was first formulated by Poincaré and was elaborated in recent years at great length by Grünbaum (1973). This view is mistaken: the isometry hypothesis is not an arbitrary convention but a well founded conjecture. Indeed, isometry is based on well confirmed physical theories that account for it. Thus the isometry of atomic time standards (the most accurate of all standards) is explained by the atomic theory. Indeed, the latter shows that the frequencies of the light waves emitted by the atoms constituting the standard depend only upon the variables that happen to be kept constant (pressure and temperature). In other words, statements such as "The atomic time standard has a constant period (at constant pressure and temperature)" are neither conventions nor raw experimental data, but well founded and well confirmed hypotheses representing certain isometric facts. The conventionalist mistake might have been avoided if standards had been carefully described and distinguished from units. Correcting that mistake destroys one of the pillars of the conventionalist philosophy of space and time, to which we shall return in Sect. 3.

The next item in our agenda is general physical theory. Every such theory poses a number of philosophical problems, some of which we have studied earlier (e.g. Bunge 1967c, 1973a, 1983b). Here we shall touch only on one of them: the difference between fundamental (or mechanismic) and phenomenological (or black box) theories. (See also Vol. 5, Ch. 9, Sect. 1.2.) *phenomenological theories*, such as the elementary theory of electrical circuits and classical thermodynamics, describe large masses of facts without explaining them, i.e. without accounting for the mechanism of their occurrence: to indulge in slang, they tell it like it is, not why it is so. Every phenomenological theory contains magnitudes that are readily accessible to measurement. (However, many of them also contain magnitudes that can be calculated but not measured in a direct fashion. For example, thermodynamics contains, among other unobservables, entropy and specific heat at constant volume.) Consequently the phenomenological theories can be confronted rather directly with laboratory data. (For this reason they are the darlings of radical empiricists, such as the classical positivists.) But the price paid for empirical testability is high. For one thing phenomenological theories are only skin deep. For another they contain parameters that do not represent any physical properties and that have to be determined experimentally.

Fundamental theories, on the other hand, contain no such parameters. All the variables and constants occurring in them represent, more or less symbolically, physical properties. For example, Dirac's relativistic quantum mechanics contains only four constants, all of them universal and with a reasonably clear physical interpretation: the values of the electron mass and charge, the velocity of light in vacuum, and Planck's constant. Yet paradoxically, although fundamental theories represent facts more directly than phenomenological theories, they are far more remote from raw laboratory data. This is due to the fact that they are not about observable facts. The function of fundamental theories is not to fit empirical data but to explain phenomenological laws and theories. For example, Newtonian mechanics explains Kepler's laws, and quantum mechanics explains phenomenological solid state theory. Because of their depth and their remoteness from raw data, fundamental theories have often been opposed by empiricists.

Another important point usually glossed over by philosophers is the difference between general and special theories. A *special theory* is obtained from a general one (when available) either by dropping or by specializing certain hypotheses. (In the first case the special theory is a subtheory of the more comprehensive one. Thus statics is a subtheory of dynamics, and electromagnetism *in vacuo* is a subtheory of electrodynamics.) For example, particle mechanics, rigid body mechanics, hydrodynamics, and celestial mechanics are so many specializations of continuum mechanics. (Every such special theory may be called a bound model; a special theory that is not based on a general one may be called a free model. See Vol. 5, Ch. 9, Sect. 1.2.) The difference is philosophically and didactically interesting for a number of reasons. For one thing the general theory cannot be obtained from the corresponding special theories by an inductive process: the general-special relation is deductive. (Yet some reputable scientists and philosophers have propagated the myth that continuum mechanics can be built out of particle mechanics.) Secondly, the more specialized a theory the better testable it is, because all test situations are particular. Moreover, all the tests of general theories are indirect, namely via special theories. (For example, the best tests of Newtonian mechanics are observations in planetary astronomy.) Thirdly, physics makes contact with engineering at the laboratory and on the level of special theories. The reason is clear: engineers are particularly interested in theories accounting for the physical aspects of artifacts – e.g. theories of the flow of liquids of certain types, or of directional antennas, or of nuclear reactors.

Finally we come to the *metatheoretical principles* in physics, such as those of correspondence, combined parity (CPT), and general covariance. (The adjective 'metaphysical' would be more precise but it has been preempted as a synonym of 'ontological'.) The metatheoretical principles are of course about theories or some of their components, such as laws; when they refer to the latter they can be called 'metanomological'. (See details in Bunge 1961.) Let us examine them quickly.

The best known of them is the correspondence principle, which was employed explicitly in building the two relativities and the quantum theories. It states that the more general theory should reduce to the more special one in the domain in which the latter holds. For example, special relativity should (and in fact does) reduce to classical physics for small velocities; and in turn it should (and in fact does) hold as a good approximation of general relativity in flat spacetime. The principle of correspondence is heuristic during the stage of theory construction; once the new theory has been built, the principle is proved with paper and pencil: it has become a metatheorem. The methodological interest of the correspondence principle is obvious: it functions as a guide for building and evaluating new theories. The mere existence of principles of this kind refutes the fashionable opinions that "anything goes" in science, that theory choice is a purely subjective affair, or that it is governed exclusively by external (social) considerations. It also refutes the view that scientific revolutions raze the past: it shows that present knowledge guides the search for new knowledge, that the march of science is evolutionary. (See Vol. 6, Ch. 12, Sect. 3.1.)

The combined parity (CPT) theorem is a metatheorem that can be proved rigorously. It states that all local (nearby action) field theories (such as classical and quantum electrodynamics) are invariant under the simultaneous reversals of all charges, parity, and time. These reversals are conceptual, not factual. Thus, the theorem tells us that, if *we* invert the charge *sign* (on paper), then *we* must also invert the parity and time *signs* (on paper) if we wish our basic field *equations* to remain invariant. Clearly, then, the theorem, unlike a conservation theorem, say, does not represent a law of nature: it is a metanomological formula. Hence it cannot be subjected to experimental tests except indirectly, i.e. through the test of some other consequences of the theory concerned. Its main philosophical interest lies perhaps in suggesting that, in principle, it is possible to build one rigorous metatheory (however small) for any given physical theory.

As for the general covariance principle, it states that the basic formulas (though not all of their consequences) of a fundamental theory ought to remain unchanged under general linear coordinate changes and, in particular, under coordinate transformations representing exchanges of reference frames (hence in particular of observers). The theories that are in fact generally covariant are naturally called 'general relativistic'. Those covariant under more special coordinate transformations receive special names, such as 'Lorentz covariant' or 'Galileo covariant'. Contemporary continuum mechanics contains a similar principle that refers not to the fundamental equations of motion or field equations, but to the constitutive equations, which describe constant relations among properties of materials of special kinds. The principle, called of 'material frame indifference' or of 'material objectivity', states that the constitutive equations must be invariant under changes of reference frame (a fortiori of observer).

These various covariance principles are philosophically interesting for various reasons. First, they are not laws of nature but *principles of theory construction which become metatheorems* once the new theory has been built. Second, although some of them are heuristically powerful, none has deductive power (i.e. nothing follows deductively from any of them). Hence they should not occur among the axioms of a theory – *pace* a number of distinguished authors. Third, they show that, although some of the individual properties of things are relative or frame-dependent, the basic laws are not: the latter are absolute or coordinate-free. A fortiori, these laws are invariant with respect to changes in observer, i.e. they are the same for all observers. Therefore the covariance principles constitute a pillar of scientific realism.

Note that physics contains metatheoretical principles of two kinds: theorems and rules. The Lorentz covariance of Maxwell's equations and the CPT invariance of local quantum field theories are mathematical truths and, more particularly, metatheorems belonging to the metatheories of the corresponding physical theories. On the other hand the correspondence principle, the covariance principles, and the quantum-mechanical rule according to which every "observable" (dynamical variable) ought to be represented by a hermitian operator, are heuristic (involved in theory construction) or methodological rules or norms. Being different they should be evaluated differently: the theorems by canons of mathematical rigor, the rules by the value of the theories they help build. Still, we have also seen that the theorem/rule distinction is historically relative, in the sense that what begins as a heuristic rule may end up by being proved as a metatheorem.

The metatheoretical principles are metastatements not object statements. Does this entail that they lack any factual reference or content? This is what Kretschmann (1917) held with regard to general covariance, though on different grounds. We proceed to prove that the metatheoretical principles of physics referring immediately to law statements refer mediately to the physical objects referred to directly by the latter; so, they do have a factual content after all. The proof employs the principle of our theory of reference

(Vol. 1, Ch. 2) according to which the reference class of a predicate equals the union of the cartesian factors of the domain of the predicate. Let P be an object predicate, i.e. one representing a property of concrete objects of some kind. For the sake of simplicity let us assume that P is a nondimensional magnitude representing an intrinsic (absolute) property of things of kind A. In short, assume that $P: A \rightarrow S$, where S is the set of statements containing P. Consider next a metapredicate M, such as "is generally covariant", or "is true". M may be construed, similarly, as a function $M: S \rightarrow SS$ from statements S to metastatements SS (containing M). Finally form the composition of the functions P and M in this order. I.e. build the function $Q = M \circ P: A \rightarrow SS$ shown in the commutative diagram

$$\begin{tikzcd}
A \arrow[r, "P"] \arrow[dr, "Q"'] & S \arrow[d, "M"] \\
 & SS
\end{tikzcd}$$

Obviously, the new predicate Q is ultimately about A's. In other words, the manifest or *immediate* referent of a metanomological statement is a law statement; but its deep or *mediate* referent is the same as that of the corresponding law statement. *Example* Law statement: "The Angular momentum of a body in a central force field is constant". Metastatement: "The preceding statement is translation invariant". Composition: "The constancy of the angular momentum of a body in a central force field does not alter under translation". This rigorous result solves Kretschmann's problem, discussed endlessly in the literature.

2. TWO CLASSICS

2.1. *Classical Mechanics*

Just as arithmetic used to be regarded as the queen of mathematics, so mechanics was the queen of physics until the birth of field theories *ca.* 1850. Mechanics is not only the first comprehensive and successful factual theory to be proposed in history: it is also the one dealing with the most conspicuous of all processes, namely the change of place of perceptible bodies. No wonder then that mechanics is presupposed by other physical theories and that it has attracted the attention of so many eminent mathematicians and philosophers. In fact, analysis was originally introduced, in part, to model motion, and classical mechanics is the daughter of ancient atomism – the first mechanistic philosophy – as well as the mother of modern mechanism. Yet, despite its antiquity, research in classical mechanics is still producing novelties: see the *Archive of Rational Mechanics and Analysis*. And the philosophy of mechanics is still in its infancy. Here we shall be able to touch on just a handful of philosophical and foundational problems in classical mechanics. (See Truesdell 1984 for more.)

To begin with, let us be clear about the objects or referents of classical mechanics: they are bodies, i.e. physical objects endowed with mass and precise (though often complex and variable) geometrical properties, such as a definite shape. Massless objects, such as photons and neutrinos, are studied by field theories; and shapeless objects, such as electrons, are studied by quantum "mechanics". Mechanics handles forces in a phenomenological manner, i.e. without assuming any field mechanisms. Consequently mechanics can describe a body as a whole and even its state of internal stress. But it cannot account for the very existence of bodies, all of which are composed by atoms held together by fields of various kinds, in particular electric fields.

In the preceding paragraph we have inadvertently and tacitly rejected two myths that are rather popular among physicists and philosophers. One of them is the belief that the mechanics of extended bodies (i.e. continuum mechanics) can somehow be reduced to the mechanics of point particles. This is of course impossible: a continuum cannot be built up from separate points. On the contrary, the mass point results from specializing the density function to a sharply peaked one (a delta). In other words, particle mechanics is a subtheory of continuum mechanics, not the basis of the latter. The second myth is that classical mechanics underlies all other theories of massive entities, in particular quantum mechanics. (This myth originates in another, namely the operationist dogma that quantum mechanics refers exclusively to experimental situations, all of which are supposed to be describable in classical terms. More on this later.) This too is a mistake, because (*a*) only quantum mechanics is in a position to explain the formation and properties of extended bodies out of smaller entities (which are not classical particles), and (*b*) it is in several ways incompatible with classical mechanics.

We have stated that mechanics is a factual science because it accounts for bodies. This is obvious from a semantic analysis of any of the predicates of mechanics, such as the concepts of mass and position of a body. However, from Kant down to the dawn of our century it was usually assumed that mechanics was a part of mathematics: that its principles were truths of reason, perhaps even necessary ones, rather than fallible hypotheses. (See e.g. Whewell 1847, Part I, p. 248.) The conventionalists went so far as to state that the axioms of mechanics were disguised definitions. Everybody called *rational mechanics* what we now call 'classical mechanics' – and all mathematics students were required to study it.

There were several reasons for the belief that mechanics is an a priori science like mathematics. One is that, from a purely formal viewpoint, every formula of mechanics is a formula of analysis: in this regard mechanics is, indeed, only a special case of mathematics. (But, of course, no formula of analysis is in fact one of mechanics unless it be interpreted in mechanical terms. Recall from Ch. 1, Sect. 4.1 the need to enrich a mathematical formalism with semantic assumptions in order to generate a factual theory.) Another reason for the belief we are commenting on was the apparent perfect fit of mechanics to analysis – which should not have been surprising given that the calculus had been invented largely as "the mathematics of motion", to the point that functions were often "defined" as the trajectories of moving points. A third root for the mistake is that, as in the case of other fundamental theories, most of the evidence for mechanics came after it was built. This too is understandable, since the very design and interpretation of any experiment involves some knowledge of mechanics. In any event, the advent of the special theory of relativity (1905) demolished the aprioristic thesis by showing that the laws of classical mechanics hold only for slow motions: they are not synthetic a priori truths.

From Kant till the birth of general relativity (1916) it was commonly believed that at least one ingredient of classical mechanics (and most other physical theories as well) was *both* factual (or synthetic) and a priori, namely geometry. This belief was perfectly reasonable in view of (*a*) the excellent fit of Euclidean geometry to physical space (in the small) and (*b*) the familiarity with the procedure of mathematizing particles as points, lines of flow as curves, bodies as regions of space, and so on. It was eventually realized that this procedure is just the dual of the procedure of interpreting geometrical objects as physical (in particular mechanical) ones. Nowadays we all understand what Gauss had suspected in 1821: namely that there is no identity between mathematical and physical geometries, but that a few geometries can be made to represent the spatial properties of physical objects and their relations. Hence, when axiomatizing classical mechanics we place Euclidean geometry in the formal background (or bag of presuppositions) of the former, and press it into service every time we need to describe the spatial properties of bodies – e.g. the boundary of a body or the shape of a flow line.

Although it is now generally understood that a physical geometry results from enriching a mathematical geometry with a set of semantic assumptions ("correspondence rules"), not all aspects of the question are clear. One of them is this: What view concerning the nature of space is classical mechanics committed to? In other words, what ontological geometry does classical mechanics presuppose? It may be remembered that there are two families of such geometries: the absolutistic and the relational ones. (Vol. 3, Ch. 6, Sect. 1.1.) According to the former space is self-existing (absolute), whereas according to the latter space is the basic network of relations among physical objects, so that space would cease to exist if it were possible for the world to be annihilated.

Now, it is usually believed that, just because Newton happened to favor absolute space and time, classical mechanics presupposes an absolutistic theory of space and time. This is mistaken: classical mechanics is consistent with either ontological geometry. (Moral: The history of science can be misleading unless accompanied by the foundations and philosophy of science.) In fact the distance functions occurring in classical mechanics are defined on pairs of spatial points that may or may not be occupied by particles; and time may but need not be regarded as a function defined on pairs of events. (See Noll, 1967.) Thus it makes no difference to mechanics whether its independent variables x, y, z, and t are regarded as real numbers taken in themselves or as values of one or more (coordinate) functions defined on domains containing sets composed of physical objects such as things or events. (In particular, the equations of motion remain the same whether t is regarded as an arbitrary real number, as the parameter of a certain continuous group of transformations, or as an arbitrary value of a time function such as, e.g., $T\colon E \times E \times F \times U_T \rightarrow \mathbb{R}$, where E is a set of events, F a collection of reference frames, and U_T the set of all possible time units.)

In short, it is not true that classical mechanics is wedded to absolute space and time: in fact it is neutral in the conflict between the absolutistic and the relational views of space and time. On the other hand experimental mechanics and, in general, experimental physics adopts unwittingly the relational view, for we can measure only distances between physical entities, and durations between events. So, if we wish theoretical physics to match experimental physics, we must adopt a relational ontological geochronometry. However, this will show only in general relativity (Sect. 3.2). Mechanics is quite insensitive to the choice between the two views on the nature of space and time.

A related topic is that of the nature of reference frames. With but a few exceptions physicists identify reference frames with coordinate systems. However, a coordinate system is a mathematical concept and therefore an object that cannot move relative to anything; on the other hand reference systems are *physical* systems that can move relative to other frames. For this reason the more careful authors state that coordinate systems map or represent (actual or possible) reference systems, which are distinguished physical systems such as nearly rigid bodies, possibly with clocks attached to them. (See Bunge 1967c.) In any event, whenever solving a particular problem in any theory containing the concepts of space and time, we must adopt both a physical reference system (e.g. a laboratory or the sun) *and* a coordinate system of some kind (e.g. cartesian or cylindrical) representing the latter.

What holds for coordinate systems holds for coordinate transformations. We must distinguish the purely mathematical relation between two different coordinate systems from a frame transformation such as a galilean one. The former relates the coordinate systems, e.g. the translation combined with dilation summarized in the formula $\ulcorner x' = ax + b \urcorner$, where a and b are real numbers. On the other hand the galilean frame transformation $\ulcorner x' = x - vt \urcorner$ relates the positions of a *physical* object relative to two different *frames* moving relative to one another at speed v, after the duration t has elapsed. Likewise the Lorentz and the general relativistic transformations concern reference frames not just coordinate systems.

The notions of reference frame and frame transformation are central to mechanics because all positions, and a fortiori all motions, are frame dependent. The relativity of motion is obvious from the fact that it can be transformed away (at least locally) by adopting a comoving reference frame. Of course, any such "rest description of motion" proves only that motion is relative, not that it is illusory – the more so since any object at rest relative to some frame moves relative to other frames.

Reference frames are so important that they are usually involved in the very formulation of the basic laws of motion. In fact, the correct formulation of these laws is as follows: "There exists (really) at least one reference frame relative to which the following equations hold [and here the equations are written]." Once the laws have been stated we can introduce the definition of a particularly important type of reference frame: "Any reference frame relative to which the preceding laws statements hold is called an *inertial* frame". Since different theories are characterized by different equations of motion, there are different types of inertial frame: more on this in Sect. 3.1.

Just as we distinguish coordinate systems from reference frames, we must not confuse the latter with observers – another common confusion in the physical and philosophical literature. We must avoid this confusion for the following reasons. Firstly, an observer is only a very special kind of reference frame. Moreover, when attached to Terra observers are not inertial frames, hence they do not serve as referents (together with bodies) of the basic laws of mechanics. Secondly, even if observers were adequate reference frames on Earth, it is unrealistic to assume that they could live everywhere in the universe. In any case most reference systems used in physics (including astronomy) are uninhabited. Thirdly, the confusion between reference frame and observer invites subjectivistic misinterpretations of physics. Thus the true statement "All wavelengths are frame-dependent" can be misinterpreted as "All wavelengths are observer-dependent, i.e. subjective, i.e. physically unreal". More on this in Sect. 3.1.

The distinctions between reference frame on the one hand, and coordinate system and observer on the other, are crucial in the formulation and discussion of the relativity principle. There are several inequivalent versions of this principle. The most popular ones concern the alleged invariance of phenomena, or even experiments, as viewed by different observers. However, not all facts are the same in (relative to) all frames of reference. For example, the Doppler effect does not occur at the light or sound source. (Nor do law statements remain formally invariant when rewritten in different coordinates; for example, when translating a wave equation from cartesian to spherical coordinates additional terms appear.) The correct version of the relativity principle is about laws, not facts, and about reference frames, not observers. It reads thus: ⌜ The *basic* law statements *ought to be* the same regardless of the reference frame⌝. This is, of course, a normative metanomological statement (Sect. 1.2.).

Let us now turn to somewhat more substantial issues, in the first place the notions of mass and force. Both have been the object of much philosophical speculation since the time of Newton. They were even the pillars of the *Kraft und Stoff* materialism popular in the 1850s. Mach and other anti-materialists tried to rid physics of the two concepts by defining them. (Mass was "defined" as relative acceleration, and force as the product of mass by acceleration.) However, we now know that such "definitions" are circular because they presuppose Newton's laws of motion, which contain the mass and force concepts. (See Bunge, 1966.) The concept of mass, which seemed obscure and suspect one century ago, is nowadays construed exactly in the same way as Newton did, namely as the quantity of matter of the kind that mechanics deals with. Thus the atomic mass of an atom is roughly equal to the number of nucleons contained in its nucleus.

So, if anything, the concept of mass is even more deeply ingrained in physics than before Mach's criticism of it. The same holds for the concept of force: instead of eliminating it, physics has multiplied (and inter-related) the number of types of force – or, more precisely, of fields. True, it is mathematically convenient to describe fields in terms of potentials rather than forces. However, given a potential the corresponding force is definable – in the simplest case – as the gradient of the potential. True, static potentials are characterizable in purely geometric terms. But in general they are also time dependent. True, for operationists and conventionalists fields are just fictions convenient to summarize certain empirical data. But we know this thesis to be false for the following reasons: (*a*) fields can propagate at measurable speeds and they can interact with chunks of massive matter, such as electrons; (*b*) all the field "quantities" (e.g. the potentials) refer to putatively real fields, this being why they are studied in physics rather than in mathematics; (*c*) sometimes fields (and their equations) are postulated, and only later on experiments to test such hypotheses are devised with the help of the corresponding field theory.

Let us now turn to a couple of metatheoretical problems. But before doing so we must characterize the notion of a well posed mechanical problem, because some metatheoretical questions are pseudoproblems arising from overlooking some of the ingredients of a *well posed* mechanical problem. These ingredients are:

- the equations of motion (or the variational principles entailing them),
- the force laws to be plugged into the equations of motion,
- the constitutive equations characterizing the particular materials concerned,
- the constraints on the position coordinates,
- the empirical data (or corresponding prescriptions): mass density (or total mass as the case may be), stress tensor, initial conditions (initial values of positions and velocities), and boundary conditions (values of certain properties on the body border).

The above items are also necessary for posing correctly certain metatheoretical problems. One of them is whether classical mechanics is time-reversible, i.e. invariant with respect to the change in the sign of t. The usual answer is affirmative because it is limited to examining the equations of motion, which, indeed, are *T*-invariant because the acceleration is, and mass and charge are tacitly assumed to be *T*-invariant as well. But this reasoning is incorrect because it overlooks the other ingredients of the problem, in particular the constraints, boundary conditions, and constitutive equations, which restrict the realistic solutions to the equations of motion. The elementary example of Figure 2.2 suffices to refute the common belief that, unlike quantum mechanics, classical mechanics is time-reversible, i.e. *T*-invariant (Truesdell 1974, 1982a).

A similar example of motion with impact can be used to explode the even more popular myth that classical mechanics is the paragon of a "deterministic" theory: Figure 2.2(*c*). A mass point strikes a wedge and is deflected either to the left or to the right. In this case there is no unique solution, so it is impossible to make an unambiguous prediction from knowledge of the equations of motion and the initial conditions. (See Truesdell 1974.) The problem calls for the assistance of probability theory. A quickly growing body of probabilistic classical mechanics (see Axelrad 1983) is the best epitaph to the myth that classical mechanics is inherently "deterministic" (non-probabilistic). What characterizes classical mechanics is not so much that it fails to contain probabilistic concepts, but that it conceives of bodies as possessing definite shapes, positions, and trajectories – hence also sharply defined velocities, angular momenta, and energies. This assumption is relaxed in quantum mechanics: See Sect. 4.

Even more striking illustrations of "indeterministic" and therefore un-

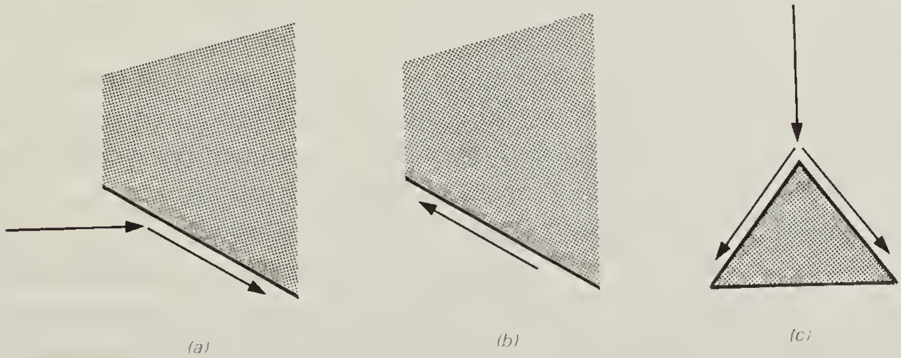


Fig. 2.2. A motion that is not invariant under time reversal, i.e. under the reversal of the velocity (and spin if any). (*a*) Original motion with impact; (*b*) reversed motion. In this case reversibility is obtained by requiring that the velocity (not just the trajectory) be continuous (Clifford Truesdell, personal communication, 1982). (*c*) A motion with indeterminate outcome: the incoming path bisects the wedge angle (Truesdell, 1974).

predictable behavior occur in non-linear classical mechanics. If the equations of motion are non-linear then the steady state orbit of the system (its "attractor") may be "chaotic" (irregular), hence unpredictable. The motions on such an attractor *look* chaotic although they satisfy ("are ruled or controlled by") nonprobabilistic ("deterministic") equations of motion. Such orbits are characteristically unstable in the sense that a minute change in the initial conditions may result in a large difference in the orbit; moreover bifurcations emerge, so that the position coordinate may take almost any value. (A well-studied example is the finite difference equation $\ulcorner x_{n+1} = \lambda x_n(1 - x_n)\urcorner$, where n is a natural number, and x ranges between 0 and 1, whereas λ can vary between 0 and 4. For a given value of λ there is a bifurcation to two different steady state solutions. As λ increases, each of these solutions branches out into another two, and so on. For a certain critical value of λ bifurcation occurs infinitely many times, i.e. x can assume any value. See e.g. Hellman, Ed. 1980.) Morals: (*a*) randomness can be simulated by perfectly "deterministic" systems, hence it cannot be diagnosed from a mere inspection of data such as observed trajectories; (*b*) determinism, an ontological category, cannot be defined in terms of predictability, an epistemological one (Bunge 1959a).

This subject dovetails with that of the next subsection.

2.2. *Statistical Mechanics*

Classical mechanics deals with bodies and systems of such. The links among the system components, such as contact and gravitational forces, generate and preserve the system itself, as each component codetermines the behavior of every other component. As soon as these links weaken, the components acquire some leeway, as a consequence of which some measure of randomness is bound to creep into the system. Thus the phases (or initial displacements) of an aggregate of mutually independent (or nearly independent) oscillators are bound to be distributed at random.

Classical mechanics is incompetent to handle systems composed of huge numbers of weakly interacting things – such as gases – except as black boxes, i.e. by deliberately ignoring their composition and structure. The study of systems composed of a very large number of units (typically 10^{27}) linked by weak forces, and a fortiori systems of mutually independent components, belongs to statistical mechanics. This theory assumes classical mechanics and it contains randomness hypotheses, such as probability distributions of initial positions and velocities. Hence it uses not only the mathematical tools of mechanics but also the probability calculus. In other words, statistical mechanics results from the merger of mechanics and certain probabilitistic hypotheses.

Determinists of the classical type regard this merger as artificial and necessitated by our ignorance of microscopic details rather than by the nature of things. In other words, they regard chance as a synonym of ignorance, hence as an epistemological category rather than an ontological one. They argue that if only we could get to know the initial position and velocity of every single atom in a gas or a liquid, we would not need the probability concept. (Put in theological terms: God does not need probability nor, a fortiori, statistical mechanics.)

This argument overlooks the fact that the independence (or near independence) of the components of a system such as a gaseous or liquid body is a source of objective chance or randomness (not of chaos or lawlessness). According to classical statistical mechanics chance is neither in the individual system components nor in the eyes of the (blind) beholder, but in the nature of things: it is an *emergent property* on the same footing as other system properties such as entropy and temperature. (Recall the definitions of emergence in Vol. 3, Ch. 2, Sect. 4.3, and Vol. 4, Ch. 1, Sect. 3.2.) The simplest randomness (or chance) hypotheses in statistical mechanics concern the initial positions and velocities of the system components. But these are not the only ones. Thus the modeling of the crystallization process in a liquid calls for probabilistic hypotheses concerning the nucleation process. Since these events are random in space and time, neither the place nor the time of the emergence of the first crystal can be predicted exactly.

A key probabilistic concept of statistical mechanics is that of thermodynamic probability, or number of microstates (of the components) compatible with a given macrostate (of the system). This is an objective property of the system, namely a measure of the play, leeway, or "degeneracy" of the system components. It is an important property because it is used to calculate another system property, namely its entropy. In fact, the simplest measure of the entropy of a system makes it proportional to the logarithm of the thermodynamic probability of the system.

The thesis that thermodynamic probabilities are *objective* system properties is confirmed by the fact that they can be measured in many cases. (In the case of quasistatic changes one measures temperatures and quantities of heat, and calculates the entropy by means of the second law of thermodynamics.) But if one interprets probability in a subjectivist manner – e.g. as a measure of one's information, or rational belief – then the entropy must be interpreted in the same manner. That is, entropy ceases to be an objective property of multicomponent systems, to become a measure of some knowing subject's uncertainty concerning the precise microphysical configuration of the system. (As one physicist put it, entropy measures "our difficulty of deciding what state the system is in".) This subjectivist interpretation turns the law of increasing entropy (or 2nd law of thermodynamics) into a law of the psychology of knowledge. Obviously, if physics is to remain an objective science – pardon the redundancy – then one will speak of the play (or "degeneracy") of microstates compatible with a given macrostate, instead of the subjective probability concerning our knowledge of such states. (More on subjective probability in Ch. 1, Sect. 4.2.)

Yet Jaynes (1979) appears to have reformulated classical statistical mechanics using the subjectivist concept of probability (as a measure of the plausibility of a hypothesis) and a subjectivist interpretation of information theory (as measuring the decrease in one's ignorance upon receiving a message). How has it been possible for him to obtain correct general results, even new ones, with such obviously incorrect interpretations? There are several reasons for this, among them the following. Firstly, the final formulas in statistical mechanics do not involve probabilities of elementary events but only macrophysical magnitudes, in particular averages; so, no subjectivist assumptions occur in them, Secondly, Jaynes has simplified matters by assigning uniform probabilities to the macrophysically irrelevant details; so, the estimates of such probabilities are not a matter of subjective judgment after all. Thirdly and most important, Jaynes's subjectivism is quite limited, for he does not claim that a thermodynamic system possesses an entropy only because we do not know (or do not care to find out) its microphysical details. Rather, he states that when using an entropy function we deliberately disregard such details. In fact he writes explicitly: "The entropy of a thermodynamic system is a measure of the degree of ignorance of a person *whose sole knowledge about its microstate consists of the values of the macroscopic quantities* X_i which define its thermodynamic state. This is a completely 'objective' quantity, in the sense that it is a function only of the X_i, and does not depend on anybody's personality. There is then no reason why it cannot be measured in the laboratory" (Jaynes, 1979, p. 42). Clearly, then, the interpretation of entropy as a measure of the degree of ignorance is an *idle assumption* that does not occur in the calculation or in the interpretation of the final results. In conclusion, Jaynes's version of statistical mechanics is just as objective as the standard one. (There are two further reasons, of a technical nature, for Jaynes's success. One is that he does not confuse probabilities with frequencies, so he has not been slowed down by having to imagine impossible experiments. The other is that he has refrained from using the controvertible ergodic hypothesis.)

So far we have been concerned tacitly with what is usually called the Boltzmann model of a complex system with more or less random components. There is an alternative model (or method) that at first sight is neither quite as "indeterministic" nor as realistic as Bolztmann's, namely Gibbs's. The Gibbs model refers not to a single system but to a whole *imaginary* ensemble of similar copies of one and the same system. Each copy differs from the others in some microphysical detail. Each copy is supposed to behave "deterministically", but the price for this apparent gain is that the probability distributions are defined over the imaginary ensemble (or collection of copies). Nor is the Gibbs model any less realistic than the Boltzmann one, for it postulates that the real aggregate behaves like the average over the given imaginary ensemble. Thus the two models represent differently one and the same real macrosystem, and both are probabilistic.

Whether in the Boltzmann or in the Gibbs formulation, statistical mechanics deals with both reversible and irreversible processes. A reversible process is one that takes the system back to its initial state without any extra energy expenditure. A convenient way of representing the reverse of a process is to invert the sign of the time variable or, what is equivalent, to reverse the velocities of all the components. For a process to be reversible it is necessary, though not sufficient, for its laws to be time-reversible (or *T*-invariant). We saw in Sect. 2.1 that *T*-invariance is only necessary for motion to be reversible. Another example is the propagation of a spherical electromagnetic wave: unless the wave is generated at the center of a perfectly reflecting spherical surface, there is no way it will contract to a point.

Statistical mechanics associates irreversibility with entropy increase without identifying them. Irreversibility is necessary but insufficient for entropy increase. In fact there are irreversible processes which are not thermodynamical. For instance, introducing a coin into a piggy bank is far easier and requires less energy than taking it out. The reason is that in the former case we control the coin and therefore the boundary conditions, whereas in the latter case we control only the system as a whole, and so must rely either on blind shaking or on brute force. Another example is the transmutation of an electron-positron pair into three gamma photons. This is usually described as a reversible process. (Moreover all elementary processes are said to be reversible.) But it is not, for the chance of three photons meeting at a given place is very slight.

Statistical mechanics was originally invented to explain thermodynamics, the paragon of a phenomenological or black box theory. Thus, statistical mechanics explicates the total energy of a system as the sum of the kinetic and potential energies of its components, and the entropy of the system as the "degeneracy" or leeway of its microstates relative to its macrostates. Thermodynamics is probably the only important physical theory to have been born in technology – but of course raised in mathematics. This explains why its principles are still often couched in terms of thermal machine cycles, or even in anthropocentric terms. Thus its most famous (yet still intriguing) principle – the second – is sometimes stated thus: "The energy of an isolated system is forever degrading into less useful forms". A more sober formulation of this principle is this: "The total entropy of an isolated system does not decrease in the course of time". (This holds only in the long run: there are temporary random fluctuations. I.e., the 2nd law is probabilistic.)

Since its inception the second principle of thermodynamics has been used to buttress the eschatological view that everything in the world is doomed to ultimate decay, to the point that the entire universe is bound to end up in a state of "thermal death" or maximum entropy. This would actually be the case if the universe were finite. But there is no firm cosmological evidence for this assumption; therefore we had better suspend judgment on the total entropy of the universe. Besides, even assuming that the universe were a closed system and therefore subject to the 2nd law, nothing would prevent it from containing subsystems the total entropy of which were constant or even in the process of decreasing. In fact this is what happens in the formation of stars, star clusters, crystals, organisms, etc. We shall return to this problem in Sect. 8.2.

Another popular myth is the belief that, since everything is fated to run downhill, the emergence of biosystems violates thermodynamics and calls for nonphysical, perhaps even theological, principles. This is a gross mistake, for we all know that biosystems are open systems and therefore not subject to the 2nd law. Every time a protein molecule is synthesized, or a cell divides, a negentropic process occurs. The organism as a whole maintains a constant level of entropy at the expense of its environment. The 2nd principle does not apply to the organism but to the supersystem composed by the organism and its environment. (For the emergence of "structure" in open systems far from equilibrium see e.g. Glansdorff and Prigogine 1971.)

Statistical mechanics was invented to explain thermodynamics, but in fact it has so far failed to do so: the reduction is a textbook myth for the edification of undergraduates and philosophers. What has been reduced is only (*a*) the kinetic theory of ideal gases (summarized in the constitutive equation $\ulcorner pV = nRT \urcorner$), and (*b*) a portion of thermo*statics*, which is the thermodynamics of systems in equilibrium. In particular there seems to be no *general and rigorous* derivation of the second law of thermodynamics, much less of the regularities of phase transitions (such as freezing). Typically, the rigorous treatments refer to spatially infinite (and sometimes one-dimensional) systems (Ruelle 1969, 1978), or to quasistatic (extremely slow) changes (Martin-Löf 1979), or to simple monoatomic gases or mixtures (Muncaster and Truesdell 1979). Even in these cases the results are restricted to the very special case of uniform density and temperature – a state in which macroprocesses, such as convection and heat transfer, do not occur. In short, the general "reductions" are sloppy, and the rigorous reductions concern only very simple and usually unrealistic cases. The general and rigorous reduction of thermodynamics to statistical mechanics is not a fact but a program (Bunge 1973a p. 190, Brush 1983 p. 261). Nor is it reasonable to expect that a full reduction is *possible*: after all, the components of a thermodynamics system are quantum-mechanical things not classical ones. The reduction, if possible, will be to quantum statistical mechanics.

(Boltzmann and Planck are usually credited with having surmounted the most severe difficulty in reducing thermodynamics to statistical mechanics, by proving the 2nd law. Actually what they did was this: (*a*) to *postulate*, not deduce, that an isolated system ends up by having its components distributed in the most homogeneous manner possible compatible with the external constraints (container shape, fields, etc.); (*b*) to *prove* that such distribution (state) is the one that can be reached in the greatest number of ways – i.e. the one characterized by maximal thermodynamic probability; (*c*) to *prove* that in any other state the entropy is less than maximal; (*d*) to *identify* the logarithm of the thermodynamic probability with the entropy of the system, thus providing a bridge (definition, not law) between thermodynamics and statistical mechanics; and (*e*) to *suggest* without proof that any state with less than maximum entropy is earlier than the latter – i.e. that entropy increases except for small random fluctuations. Unfortunately this monumental work does not add up to a general and rigorous proof of the 2nd law, let alone of the deduction of thermodynamics from statistical mechanics – except in the imagination of elementary textbook authors. For more on this subject see Krüger 1976, Brush 1983, Truesdell, 1984.)

We close this section with comments on a couple of common mistakes. One of them is the extrapolation of the specific concepts of thermodynamics to microphysics. Thus it has become commonplace to speak of the temperature and entropy of systems of spinning "particles" such as atomic nuclei (Ramsey 1956). The price paid for this extrapolation is quite high: (*a*) one and the same metallic sample is attributed two different temperatures near the absolute zero of temperature: those of the atomic nuclei and of the rest; (*b*) negative absolute "temperatures" are introduced, which are interpreted as indicating that a system possessing any of them, far from being supercooled, is "transfinitely hot" – whatever this may mean. Similar extrapolations, at comparable prices, might be conconcted – e.g. attributing vorticity and refractive index to electrons. Moral: we must recognize the emergence of new properties and the concomitant formation of new levels. This recognition calls for the introduction of some specific concepts representing the emergent properties. In other words, we cannot expect microthings to be accounted for by macrophysical theories, or macrothings to be accounted for by microphysical theories. All we can hope for is interconnecting such theories and partially reducing some of them to others. (For the concept of partial reduction see Vol. 6, Ch. 10, Sect. 3.1.)

Our second comment refers to a philosophical distortion of thermodynamics. Because temperatures can be *measured* only in conditions of thermal equilibrium, operationists have concluded that they can be *defined* only for systems in equilibrium, so that it makes no sense to attribute temperatures to systems in states far from equilibrium, such as a rapidly cooling body. This is like claiming that, since the weight of a body can be measured only when the body is at rest on a scale, moving bodies are weightless. If physicists had admitted that operationist restriction, thermo*dynamics* would not exist, and we would know only thermostatics. In particular, the study of heat transfer, e.g. along a metal bar, would have to be proscribed, and Fourier's law declared meaningless. This is not a joke: it may well be that the operationist's concern for measurement – indispensable in the laboratory but out of place in connection with general theories – has effectively blocked the progress of thermodynamics by concentrating the attention of physicists on equilibrium states and reversible processes. (See Giles 1964 for a gallant but, alas, unsuccessful attempt to construct thermodynamics out of elementary experiences. And see Truesdell 1982b for "the disastrous effects of experiment upon the early development of thermodynamics".)

3. TWO RELATIVITIES

3.1. *Special Relativity*

Few scientific theories have been as badly misinterpreted as special relativity (SR). Let us examine briefly some of the most popular misinterpretations. (See Bunge 1967c and 1979b for more.) The earliest misinterpretation of SR was that it supports philosophical relativism, or subjectivism, for showing that a number of important physical properties, such as duration, length, and mass, are frame-dependent instead of being absolute as postulated in classical physics. The misinterpretation results from mistaking reference frames for observers (knowing subjects). Hence expressions such as 'apparent length', 'event x as seen by observer y', and even 'process x as seen by atom y', are still current in the physical literature. But reference frames, which are even more conspicuous in SR than in classical mechanics, are not observers: they are physical systems (e.g. rigid bodies) representable by coordinate systems. (See Bunge 1967c.) Moreover the only reference frame known to be inhabited by competent observers, namely our own planet, is inadequate for formulating the basic laws because it is accelerated.

Reference frames must then be distinguished from observers as well as from coordinate systems. (Recall Sect. 2.1.) In turn, neither reference frames nor coordinate systems should be mistaken for measuring devices, let alone for observers. See Table 2.1. The conflation of two or more of these

TABLE 2.1. Four different concepts often confused in the literature.

Concept	Rough characterization	Studied in
Coordinate system	n-tuple of functions (or of their values) serving to locate or label points or events in a space	Mathematics
Reference system	Macrophysical system with preferred spatial directions and possibly a clock	Physics
Measurement apparatus	Macrophysical system designed to detect events of some kind and measure some of their properties	Physics
Observer	Sentient being trained to make scientific observations of some kind	Psychology, epistemology, methodology

four concepts may produce the illusion that a mathematical definition of a coordinate system characterizes a measurement apparatus, or that every theory containing frame-dependent quantities is subjectivistic.

In classical mechanics lengths, durations and masses are absolute, i.e. the same in (relative to) all reference frames. On the other hand in relativistic mechanics lengths, durations and masses are relative to the reference frame. Therefore every body has infinitely many sizes and masses, and every process has infinitely many durations: as many as possible inequivalent reference frames. However, not all properties are relative in SR: only some classical absolutes, such as duration and mass, have been relativized; others, such as electric charge and entropy, continue to be absolute (frame invariant). In return for the loss of certain absolutes, SR has introduced absolutes unknown to classical physics, such as proper time and spatiotemporal interval. On the whole, if we count both the fundamental (undefined) and the derived (defined) magnitudes, it turns out that relativistic physics has just as many absolutes as classical physics. Indeed both sets are denumerably infinite, and each of them splits into two disjoint subsets: that of relative and that of absolute magnitudes, both subsets being infinite. (Just think, say, of the time derivatives of a position coordinate, or of the powers of the proper time.) So, in the end SR is not more relativistic than classical physics – and neither supports subjectivism.

Another popular belief is that in SR there are no privileged reference frames. (This negative statement sometimes passes for the principle of relativity.) That is not true: in SR the frames relative to which Maxwell's electrodynamics hold *are* privileged over all others. We call them *Maxwellian* inertial frames. Two such frames are equivalent if an arbitrary light ray propagates in a vacuum along straight lines with the same constant speed relative to both frames. And any two such frames are related by a Lorentz transformation (which leaves Maxwell's electromagnetic equations unchanged).

The Maxwellian inertial frames are not the same as those relative to which the equations of motion of classical mechanics are supposed to hold. We call the latter *Galilean* inertial frames. Two such frames are equivalent if an arbitrary particle (not a light ray) moves freely along straight lines with the same constant speed relative to both frames. Any two such frames are related by a Galileo transformation (with leaves Newton's mechanical equations unchanged).

The concepts of Galilean and Maxwellian inertial reference frames enable us to state one *principle of relativity* for classical mechanics and another for classical electrodynamics. Both can be compressed into a single formula thanks to the ambiguity of the word 'inertial frame': "The basic (or fundamental) laws of physics ought to be the same in (relative to) all inertial reference frames". In Sect. 1.1 we noted that the relativity (or covariance) principles refer directly to basic (not derived) law statements, and only indirectly to facts; and that they are normative or prescriptive rather than descriptive. Now we stress that there are two mutually incompatible relativity principles in classical physics: one for mechanics (hence indirectly for bodies), the other for electrodynamics (hence indirectly for fields). In other words, whereas classical mechanics satisfies the Galilean principle, electrodynamics satisfies the Maxwellian one. The two theories are then mutually incompatible. Hence they cannot be merged into a consistent theory accounting for the motion of bodies in a electromagnetic field.

Lorentz and Poincaré attempted to surmount this contradiction by keeping the equations of motion of classical mechanics and reinterpreting them in terms of absolute length contractions and time dilations caused by the motion of bodies through the aether at rest. Einstein understood that the situation called for radical surgery: it was necessary to choose between the two theories, or else reject them both. He opted for classical electrodynamics not only because it was the better confirmed of the two but also because he believed the world to be made up basically of fields. That is, Einstein assumed that classical mechanics was false and set himself the task of building a new mechanics compatible with classical electrodynamics, i.e. one the basic equations of which would hold relative to Maxwellian reference frames (i.e. be Lorentz covariant). Thus was SR born: from an examination of theories, not from an analysis of experiments. (For a refutation of the textbook myth that Einstein merely mathematized the negative experimental results of the Michelson and Morley measurements, see Shankland 1963 and Holton 1973.) The reformation of mechanics was only the first step, and one that is still not fully completed. In time the whole of classical physics will be reformed to comply with the requirement of Lorentz covariance.

What is SR about: what are its referents? Eminent physicists have claimed that it is about yardsticks and clocks, or even about measurements performed by observers moving relative to one another. This operationist interpretation of SR can be refuted in either of two ways: by subjecting SR to a semantic analysis or by axiomatizing it. In either case it is shown that the concepts of observer, measuring instrument, and measurement, fail to occur among the concepts of the basic theory. (See Bunge 1967c.) Moreover the theory is needed to elucidate these concepts as well as to design measurements inconceivable on classical physics, such as those of time "dilation". The genuine referents of a relativistic theory are physical entities. Among these, reference frames and electromagnetic (in general, massless) fields occur, as is obvious from the conspicuous occurrence of the limit speed *c*. For example, relativistic thermodynamics refers, like classical thermodynamics, to complex material systems; but, in addition, it refers to reference frames and electromagnetic fields. And the totality of relativistic physics – which is the union of all Lorentz covariant theories – refers to physical entities in general.

Another popular belief is that SR has geometrized physics and therefore eliminated becoming. This belief seems to stem from Henri Bergson, it was popularized by Hermann Weyl (1950) and was adopted uncritically by a number of philosophers of science. It is adequately described in a science fiction novel: "All moments of time exist together. The world can be thought of as a map, not only spatially, but also with respect to time. The map stretches away both into the past and into the future. There is not such thing as "waiting" for the future. It is already there on the map" (F. and G. Hoyle 1963 p. v).

Now, it is true that the history of any thing can be displayed on a spacetime diagram. But this can also be done in classical physics, and in either case it may involve an extrapolation – to the past and the future – of the observed segment of the history. If anything, becoming is even more deeply entrenched in SR than in classical physics because, according to the former, there are no instantaneous actions at a distance: all processes take some time. (SR cannot even be stated without the notion of a signal propagating with the limiting velocity *c*; a signal is a process, not a static thing.) Nor is time in SR one more coordinate, hence a purely geometric object: time is merely being represented by the fourth coordinate. The welding of space and time in SR does not rob time of its privilege, for (*a*) the concepts of space and time are mutually independent (logically irreducible), and (*b*) the changes of state of physical systems are described by formulas wherein time (or proper time) occurs as an independent variable. Spacetime has a fibrous structure, for any point in it representing a physical entity has its own light cone associated with it. And every light cone effects (*a*) an absolute (frame-independent) separation between timelike and spacelike vectors, and (*b*) a relative (frame-dependent) separation between past and future. So, the events in any given thing are not given once and for all: they happen sequentially and are thus partially ordered. (Moreover, this partial order of events implies the Lorentz group: Zeeman 1964. True, one can proceed the other way around and base SR on the group-theoretic properties of spacetime: Lévy-Leblond 1976. However, this procedure does not explain why one should adopt the relativistic spacetime metric. Worse, since it takes spacetime as existing by itself – i.e. absolutely – this procedure gives the impression that SR is a purely mathematical theory unconcerned with physical entities.)

The concept of becoming is related to those of determinacy and predictability. Positivists equate determinacy with predictability just because the latter is a symptom or indicator of the former. The equation is incorrect for the following reasons. First, determinacy is an ontological category whereas predictability is an epistemological one (Bunge 1959a Ch. 12). Event *A* may determine event *B* even though we may be unable to predict *B* from a knowledge of *A*. Second, an unpredictable event may determine a knowable one even if the chain between the two events is not stochastic. For example, Figure 2.3 shows event *A*, which influences or even causes event *B*, but lies outside the light cone of observer at *O*, who therefore cannot have knowledge of *A*, hence cannot predict *B*.

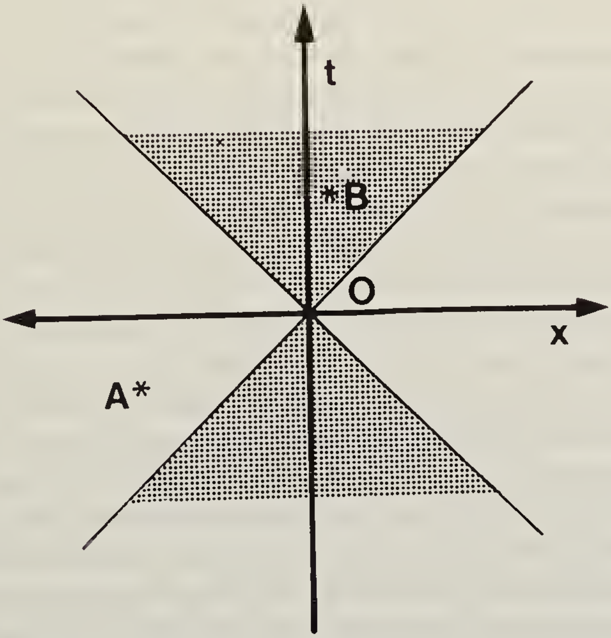


Fig. 2.3. Light emission at *A*, unknowable in principle to observer at *O*, influences event *B*, which the observer can perceive but not predict.

It is well known that SR has restricted the field of the causal relation to pairs of events connectible by light signals. What is far less well known is an even deeper conceptual modification introduced by SR in our causal thinking, namely the idea that *not every difference is an effect*, i.e. an outcome of a causal action. In fact the differences between the lengths (or durations, masses, temperatures, etc.) relative to mutually inequivalent reference frames are just differences, not effects of some events: they parallel the differences in kinetic energy relative to different frames. In particular the length "contractions" and time "dilations" are not caused by anything, although they were initially explained as effects of the aether pressure. (There is a constant output of ingenious non-relativistic mechanisms to explain such differences in causal terms. See e.g. Edmonds, 1978.) The relativistic explanation is simply that any given body has as many sizes as there are inequivalent reference frames: size is not an intrinsic property of bodies. Duration and the other variant properties are parallel: they are all relational properties. This is paradoxical only if one sticks to classical mechanics.

The last myth we wish to explode concerns the status of the simultaneity of distant events: is it fact or convention? Einstein himself, whether by a careless use of the words 'definition' and 'conventional' or under the influence of Poincaré's conventionalism, asserted initially that distant simultaneity is conventional: i.e. that, whether two distant events are judged to be simultaneous or not, is a matter of convention not of fact. Grünbaum (1973) has held that SR is *based* on such convention rather than on a set of postulates representing natural patterns – such as, e.g., the constancy (or frame independence) of the speed of light in a vacuum. The polemics unleashed by this thesis have filled too many pages over the past two decades, although the question can be decided in one minute.

In fact, in order to find out whether two distant events are simultaneous relative to a given reference frame we calculate or measure distance and travel times of light signals using one of the basic assumptions of SR. This is the hypothesis that light propagates in a vacuum, relative to any reference frame, at a constant speed c. Example: a photocell detects, at time t_2, photons emitted at time t_1 by two light flashes equidistant from the receptor and at a mutual distance d: see Figure 2.4. This is not a convention but a physical law. Moreover it presupposes a refutable hypothesis, namely that space is homogeneous and isotropic. It might well happen that the speed of light had local fluctuations, and that in some regions the speed in one direction differed from that in the opposite direction. (In this case the

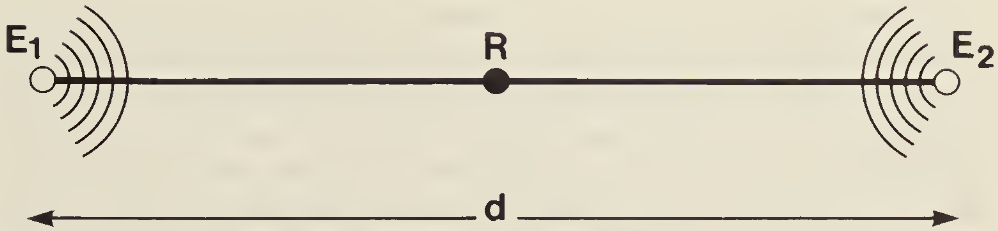


Fig. 2.4. Emitters E_1 and E_2, equidistant from receiver R, emit light pulses at time t_1 which receiver R detects at time t_2. The relation between emission and detection times is $t_1 = t_2 - d/2c$, where c is the speed of light. The simultaneity of the two emission and the two reception events, relative to the frame attached to R, is a fact, not a convention. The relation between the emission and detection times is a law not a definition.

Lorentz transformations would hold only on the average.) In short the simultaneity relation is frame dependent but not conventional: two events that are simultaneous relative to a given frame may cease to be simultaneous relative to a different frame. (In other words the simultaneity relation is ternary not binary.) And this is a trait of nature not a human convention. (More in Bunge 1968 and Torretti 1983.)

3.2. *General Relativity*

General relativity too poses a number of philosophical problems, in the first place that of the reference class of the theory. The operationist view is of course that general relativity is about observers engaged in measuring distances and durations. This statement is shown false by examining the basic gravitational field equations. These relate the gravitational potentials (represented by the space-time metric tensor) to the densities and stresses of the sources of the gravitational field, e.g. bodies and non-gravitational fields. (Briefly: $G = kT$.) General relativity refers then to matter in general and to the gravitational fields generated by matter. To be sure it also refers, in particular, to the special physical systems called ‘reference frames’; but, as we saw in Sect. 2.1, frames are not to be mistaken for observers. And, sure enough, the theory may be applied to design measurements (of e.g. the sides and angles of an astronomical triangle formed by radio signals); but the theory is so general that it does not refer specifically to any measurements. In short, general relativity does not tally with operationism – as Bridgman (1949) himself admitted. This conclusion can be proved rigorously by applying our theory of reference (Vol. 1, Ch. 2) to the basic concepts of general relativity (as dug up in the axiomatization proposed by Bunge 1967c).

Another philosophy that fails to account for general relativity is *conventionalism*, which in this case boils down to the thesis that the space-time metric is conventional (Grünbaum 1973). One argument for this thesis would seem to be that the choice of geometrical coordinates is conventional: space-time has no coordinates stamped on it, these are constructs. True, but this holds for any theory containing the concepts of space and time. Space-time would in fact be metrically amorphous only if the assignment of spatio-temporal distances were arbitrary. But it is not, except with regard to the length and time units. On the contrary, the space-time interval between two events remains constant under a general coordinate transformation. In other words, it is perfectly objective. The line element in a Riemann four-manifold is given by $\mathrm{d}s^2 = g_{ab}\mathrm{d}x^a\mathrm{d}x^b$, where $a, b = 0, 1, 2, 3$. When a general coordinate transformation is performed, the changes in the $\mathrm{d}x^a$ are compensated for by changes in the components of the metric tensor g_{ab} in such a manner that $\mathrm{d}s^2$ remains unchanged. In other words, the value of $\mathrm{d}s^2$ does not depend upon our particular choice of coordinates: it is *absolute*. This is best seen by using intrinsic (synthetic or coordinate-free) geometry.

A second argument against the conventionality (or amorphousness) of the space-time metric is this. The components g_{ab} of the metric tensor have a definite physical meaning in general relativity which they lack in the classical theory of gravitation: in fact they are the potentials of the gravitational field. (The field strengths are represented by the affinity definable in terms of the metric tensor.) A third argument is that the choice of coordinate system (hence of metric tensor) is not totally arbitrary but, on the contrary, it is restricted by the condition that the transformed time (space) coordinate be another time (space) coordinate, so that the conversion of time into space or conversely be precluded. More precisely, any coordinate systems that do not allow a distinction between space-like and time-like vectors, and all those coordinate transformations that do not preserve this dichotomy, are excluded explicitly by hypothesis. (There are further restrictions on the choice of coordinates.) Fourthly, it is true that the values of the components of the metric tensor at any point depend both upon the physical reference frame and the coordinate type. But this holds for all tensors (and vectors). What is important is that, once a reference frame and a coordinate system have been adopted, those values are not assigned arbitrarily or a priori. On the contrary, they result, via the field equations, from the distribution of bodies and non-gravitational fields in the vicinity of the point concerned. In other words, the geometric (or rather gravitational) tensor G_{ab} is determined by the matter tensor T_{ab} according to the field equations $\ulcorner G_{ab} = kT_{ab} \urcorner$, where k is a dimensional constant. This point leads to two more questions of philosophical interest.

A first question is whether general relativity effects the geometrization of physics, as has often been claimed. If this claim were true, it would refute our dichotomy of the sciences into formal and factual (Ch. 1, Sect. 1.1). But it is not. The truth is not that the gravitational field has been *replaced* by a geometric object (the geometric tensor), but that the latter has been made to *represent* the former. More precisely, in general relativity gravity is represented by space-time curvature. Thus the metric of space-time, far from being a priori and independent of the furniture of the world, is a posteriori because it depends on the content of the universe. This is obvious from an examination of the basic field equations, which, as we saw a moment ago, refer to matter of all kinds and the associated gravitational fields.

A second question is this: What happens if the matter tensor T_{ab} occurring in the field equations vanishes everywhere and at all times? In this case the solutions of the resulting equations, namely $\ulcorner G_{ab} = 0 \urcorner$, still describe, in general, a curved (Riemannian) space-time. This would seem to contradict the statement that the gravitational field causes deviations of space-time with respect to the flat (or tangent) space-time. In fact there is no contradiction, for general relativity does not state that every deviation from flatness is due to gravitation, but the converse, namely that, if there is gravitation, then space-time is curved. So, space-time curvature is an *ambiguous indicator* of gravity.

It has been claimed that the existence of solutions of the field equations for a vanishing matter tensor countenances an absolute view of space-time, i.e. one according to which space-time is the self-existing container of every thing. It is trivially true that the equations $\ulcorner G_{ab} = 0 \urcorner$ define a *mathematical* space, or rather a family of spaces. But this does not prove that these spaces are physically meaningful, i.e. that physical space-time would continue to exist even in the absence of matter. To solve this problem we need more than mathematics: we also need a theory of reference, a precise notion of physical existence, and a methodology. We have supplied these tools in previous volumes of this Treatise; let us now apply them to the problem at hand.

If the matter tensor vanishes everywhere and at all times, then $G_{ab} = 0$, and there are no *physical* properties left either. Consequently the state function of the "universe" vanishes, and correspondingly its state space shrinks to nothingness – which is precisely a mark of constructs, not of things. (Recall Ch. 1, Sect. 2.1.) In other words, space-time is in no physical state, and a fortiori it cannot jump to a different state: there are no pure spatiotemporal events, but only events occurring in physical entities such as atoms and stars. Again: nothing would ever happen in a hollow "universe" – a clear sign of its unphysical (mathematical) nature. The components of the metric tensor g_{ab}, which occur in the geometric tensor G_{ab}, cannot be interpreted as gravitational potentials in the absence of field sources for, in this case, there are no fields and consequently no gravitational potentials either. Consequently space-time cannot be adequately described by an absolutist theory consistent with general relativity. A relational theory of space-time is therefore called for. (See Vol. 3, Ch. 6 for one.)

There is, finally, the methodological problem of supplying evidence for an absolutist theory of space-time. Both theory and experiment tell us that it is impossible to shield space-time, as if it were a charged body, in order to find out what its intrinsic properties are. And common sense tells us that, since $\ulcorner T_{ab} = 0$ everywhere and at all times$\urcorner$ writes off all existents, it is incompatible with the existence of reference frames and measuring instruments. Even the weaker condition that the matter tensor vanish everywhere within a finite region of space-time is incompatible with the execution of space-time measurements in that region. If space-time were regarded as an absolute existent, rather than as a network of relations among things, its interactions with material entities, in particular measuring instruments, would have to be examined. But in this case one would not know how to proceed. In fact, a thing b acts upon a thing a if, and only if, some of the states of a in the presence of b differ from those of a in the absence of couplings with b. (Recall Vol. 3, Ch. 5.) Since space-time is in no state, we would not even know how to describe its interactions with things proper, for such interactions should cause state transitions. In conclusion, the absolutist views of space-time are methodologically as well as semantically untenable. Shorter: space and time do not exist by themselves.

The task of checking empirically the geometrical assumptions of general relativity is technically complicated and philosophically interesting. There are two methods: the direct and the indirect. The former consists in measuring some of the geometrical properties of a physical system such as a triangle composed by three electromagnetic signals. This was first done in 1964 using the Moon and two points on Earth, and its result confirmed the hypothesis that such a triangle is roughly spherical rather than plane. Similar measurements, using Mars, have corroborated that finding. However, results such as these are global: they do not allow one to determine the components of the metric tensor. To do this one must assume general relativity itself – much as any test of mechanics uses some principles of mechanics. In fact what must be done to measure the g_{ab} is to solve a system of differential equations contained in general relativity, and plug into the solutions the values, found empirically, of certain parameters occurring in the solutions (Levi-Civita 1929.)

The indirect tests of general relativity are the most numerous of all and they consist in checking the "effects" predicted by the theory, i.e. some of its theorems describing the motion of bodies or the propagation of fields. A recent test (1979) was made with the help of a Viking spacecraft lander placed on Mars, which emitted an electromagnetic signal that grazed the sun. The signal reached our planet with an additional delay of 250 microseconds, which was explained as due to the decelerating effects of the solar gravitational field.

The interest of measurements of either of the two types for philosophy is double. Firstly, they show that the determination of the space-time metric is not done by transporting yardsticks – as Grünbaum (1973) and other philosophers believe – but by making time measurements. Secondly, they show that any detailed exploration of space-time is guided by general relativity rather than proceeding in a purely empirical fashion.

General relativity poses a great many additional methodological, foundational and philosophical problems. (See e.g. Bunge 1967c, Angel 1980, Torretti 1983, Friedmann 1983.) However, we must move on. We shall now leave the neat classical picture of the world as a system composed of entities – bodies and fields – endowed with properties every one of which has, at a given instant and relative to a given reference frame, a sharp or definite value.

4. QUANTONS

4.1. *Classons and Quantons*

We now enter the wondrous quantum world or, rather, the quantum-theoretical picture of the physical aspect or level of the world. This is still a world of material things in the large sense of 'material', which includes fields; and it is just as real, if experimentally more elusive, as the world studied by classical physics. However, the things to which quantum theory refers are quite anomalous from a classical point of view. They are neither particles nor fields; they have neither definite trajectories nor precise shapes; their dynamical properties have seldom sharp values; once they clump into a system it may be hard to separate them; and their behavior, though lawful, is basically probabilistic.

We shall call *classons* the entities accounted for satisfactorily by classical physics (CP), *quantons* those that call for quantum physics (QP), and *semiquantons* (or *semiclassons*) those requiring theories that contain both classical and quantal concepts. All classons and semiquantons are ultimately composed of quantons. For example, the microscopic fibres composing this sheet of paper are classons composed of comparatively small molecules that can be described correctly only with the help of quantum theory. And the proteins and DNA molecules, which appear to be semiquantons, are likewise ultimately composed of quantons. However, "classon" is not the same as "macrophysical", and "quanton" does not coincide with "microphysical". In fact, some classons – such as dust particles – and semiquantons – such as biomolecules – are extremely small compared with perceptible bodies. And there are macrophysical quantons, such as black bodies, superconductors, superfluids, and neutron stars. (Moreover, if quantum cosmology is right, the entire universe before the big bang is likely to have been a single quanton.) Therefore our classon-semiquanton-quanton trichotomy is theoretically more profound and adequate than the usual micro-macro distinction, which is anthropocentric and therefore only practically useful.

There is no consensus about the relations between classical physics (CP) and quantum physics (QP). Indeed at least five different views have been discussed in the literature. A first view is that QP is merely an *extension* of CP, so that any physical problem, even the one posed by the collision of two trucks, could be treated quantum-theoretically if only one had enough patience and computer hours. (See e.g. Pauli *apud* Born 1971 p. 223.) This view is false because most of the problems in CP involve formulas containing concepts, such as those of shape, elasticity, and viscosity, that do not occur in the basic equations of QP. A second view is that CP is the *limit* of QP (e.g. for large quantum numbers or when Planck's constant is negligible). This is a half truth, for quantum mechanics includes parts of classical particle mechanics, and quantum electrodynamics parts of classical electrodynamics, on the average. However, QP has so far not been shown to contain hydrodynamics, the theory of elasticity, or thermodynamics. A third view is that CP is *included* in QP, which would result from enriching

CP with certain additional assumptions such as quantization rules. This influential view (Bohr 1949, Bohm 1951, Landau and Lifshitz 1958) is wrong because, in fact, QP contradicts parts of CP, and because QP entails parts of CP. (The motivation for this view is the operationist philosophy according to which every formula must be read in terms of possible experimental situations, which involve macrophysical systems, namely pieces of apparatus, that are allegedly describable in purely classical terms.) A fourth view is that QP and CP are *mutually incomparable* ("incommensurable"), so that neither reduces to the other even partially. Like the revolution operated by Galilei and Newton, which buried Aristotelian physics, QP would have buried CP. We have criticized this catastrophic view of scientific revolutions in Vol. 6, Ch. 13, Sect. 3.2. Suffice it to note here that (*a*) QP recaptures some classical results and (*b*) QP is preferred over CP in a number of research fields because, on comparing their respective results, those of QP have been found truer than those of CP. Our own or fifth view is that QP entails *some* CP in a modified manner (namely as average); that it is *very novel yet comparable* with CP; and that CP was *heuristically* instrumental in building QP, and is still indispensable to put QP to the *test*, but it is no part of the latter. (All of this can be proved rigorously only by axiomatizing quantum theory: Bunge 1967c.)

There can be no doubt that quantum theory is a good approximation to the truth – which is not to say that it is perfect. Thousands upon thousands of observations and experiments have confirmed its predictions in an amazing range of fields, from particle and atomic physics to solid state physics and astrophysics, usually with an astounding accuracy. Still, since its inception quantum theory has been the subject of vehement controversies that, far from abating as the theory became consolidated, have increased in recent years. (See Jammer 1974.) These controversies fall into two categories: *internal* and *external.* The internal controversies presuppose that quantum theory is substantially correct, and they bear mainly on the physical interpretation of its mathematical formalism. (The formalism itself is far from perfect in the case of quantum electrodynamics, noted for its trickeries, but this is a technical matter of no great philosophical interest.) The external controversies originate in a dissatisfaction with the theory as a whole, and they concern the possibility of replacing it with a classical or semiclassical theory. And the controversies of both kinds are largely motivated by philosophical reasons (or unreasons).

From a philosophical viewpoint most of the internal controversies revolve around the issue of realism, i.e. the problem of whether quantum theory supplies a realistic representation of nature or, on the contrary, is centered on the knowing subject (the observer). The internal controversies are then substantially of an epistemological kind. And most of the external controversies revolve around the issue of hidden variables, or functions of a classical type, such as classical position and momentum, which at all times have sharp values instead of probability distributions. The external controversies are then substantially of an ontological kind. Together, the controversies of both kinds span a good deal of epistemology and ontology. There is also a marginal controversy over the kind of logic suitable for quantum theory, but we shall argue that this is a pseudoproblem.

Unfortunately the two main controversies, those over realism and determinism (or hidden variables), have often been mixed up – and this by scientists of the stature of Einstein and de Broglie, Bohm and d'Espagnat. Yet the two issues are quite different: whereas the problem of realism is epistemological, that of hidden variables is ontological. (See Bunge 1979c.) Therefore one can be a realist and either admit quantum theory or hope for a hidden variables theory. Or one can be a subjectivist while accepting or rejecting quantum theory. All four stands are consistent. As for the inconsistent positions, they are probably the most numerous and popular, here as in politics.

Since we hold that quantum theory, though perfectible like any other human creation, is substantially correct, we shall engage mainly in internal criticism. Our central target will be the standard or *Copenhagen interpretation* of the mathematical formalism of the theory. A major tenet of this interpretation is that it is meaningless to speak of quantons in themselves, i.e. separately from the knowing subject: that every formula of quantum theory ought to be read in terms of laboratory operations (Bohr 1934, 1935, 1936, 1949, 1958, Frank 1938, 1946, Heisenberg 1958, Pauli 1961, Rosenfeld 1953, 1961). Thus Bohr has been quoted as stating that "There is no quantum world. There is only an abstract quantum physical description. It is wrong to think that the task of physics is to find out how nature *is*. Physics concerns what we can say about nature" (Petersen 1963). If this were true, the scientific realism upheld in Vols. 5 and 6 of this *Treatise* would be false. But we shall argue that the operationist tenet is false, and shall do so by examining the basic ideas of quantum mechanics as well as some experiments relevant to it.

Now, criticism may be destructive or constructive. If the former, it may be interesting and suggestive, but it is unlikely to dislodge any received view. To be effective, criticism must be constructive: it must offer a viable alternative. We shall therefore disregard the criticisms of the Copenhagen school made from a classical viewpoint, and shall have little patience with the attempts to "classicize" quantum theory by introducing "hidden" variables, i.e. functions that have definite ("sharp") values at all times. The nostalgic dream of a return to classical physics is just a dream, particularly after the experimental refutation of the entire family of hidden variables theories (Aspect *et al.* 1981, 1982), which will be examined in Sect. 6.2.

A constructive criticism of quantum theory can be either *reformist* or *revolutionary*: it may attempt to improve the theory or replace it with a radically new one, even farther removed from classical physics. We take it that a revolution is unnecessary at the moment, at least with respect to quantum mechanics, for there are no unavoidable inconsistencies or major experimental facts that escape the theory. (Quantum field theory is a different matter: here the problem of infinities is still to be solved in a rigorous manner.) So, reform seems to be the only reasonable course at this time. Now, reform may bear on the mathematical formalism (as is the case with every theory) or on the semantic assumptions ("correspondence rules") that confer a physical content upon the formalism. We shall deal here only with the latter, which we regard as the more urgent task and the reason that Nobeal laureate Richard Feynman has admitted that "no one understands quantum mechanics". We submit that a *realistic reinterpretation* of standard quantum theory solves its "mysteries" – without however restoring the pictorial quality of some parts of classical physics. So, ours will be a constructive reformist criticism of the Copenhagen school. (See also Bunge 1959b, 1967c, and 1973a.)

4.2. *The State Function and its Referents*

The very first question we should ask about any theory is: *What is it about*? The founders of quantum theory have proposed the most bizarre answers to this basic question. Thus Bohr (1958 p. 80 and *passim*) insisted that quantum theory refers only to "observations obtained under experimental conditions described by simple [classical] physical concepts". Heisenberg (1958 p. 52) stated that "the term 'happens' is restricted to the observation", and was of the opinion that quantum theory does not refer to nature but to our knowledge of the latter – whence it would be part of epistemology. And, in line with these opinions, Pauli (1961) claimed that the concept of a physical object independent of the observer is alien to quantum theory. It is not too surprising that such views were uttered and that they continue to be repeated in the literature: after all, quantum theory is still comparatively young and it was invented at a time when operationism was the dominant philosophy of physics. What is remarkable, perhaps unique in the history of science, is that such views, worthy of Bishop Berkeley, were not expressed by contemporary philosophers but by outstanding physicists. What is even more remarkable is that they continue to be held in a dogmatic fashion, i.e. without any critical analysis. And what is downright scandalous is that cogent technical criticisms of such dogmas are rarely accepted by leading physics journals.

What can the philosopher do when faced with amazing opinions upheld by the greatest scientists of the day? If he is tender-minded he will adopt them to buttress some philosophical school. But if he is tough-minded he must subject them to critical examination, confronting them not with alternative dogmas – the way a schoolman would do – but with facts such as the activity of quantum theorists and experimentalists. That is, he must ask himself whether those opinions match the theory as well as the experiments in which the theory is involved. Let us do just this, starting with a search for the referents of quantum theory; matters of experiment will be deferred to Sects. 5.2 and 6.1. For the sake of simplicity we shall confine most of our discussion to non-relativistic quantum mechanics in its Schrödinger (or state space) formulation.

The search for the referents of a theory can be restricted to a semantic analysis of the basic or undefined concepts of the theory. Since such concepts are bound to occur in the axioms of the theory, we may focus our investigation on them. Here we shall examine the four most important postulates. (For an axiomatization of quantum mechanics see Bunge 1967c. For discussions on the interpretations of the mathematical formalism see Bunge, 1959b, 1967c, 1973a, 1983c.)

Let us start with the central postulate: *Schrödinger's equation*, which represents the evolution in time of the state function (or vector) ψ as determined by the total energy operator, or hamiltonian $\hat{H}$:

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi \tag{1}$$

where i is the imaginary unit, $\hbar$ Planck's constant, and t time. $\hat{H}$ is a function of the position and momentum coordinates of the quanton as well as of the external field, if any. In the absence of the latter, the hamiltonian reduces to $\hat{H} = \hat{p}^2/2m$ for a single "particle", and to

$$\hat{H} = \hat{p}_1^2/2m_1 + \hat{p}_2^2/2m_2 + H_{12} \tag{2}$$

for two entities, where $\hat{p}_1$ and $\hat{p}_2$ are the momentum operators, m_1 and m_2 the masses, and H_{12} the function or operator representing the interaction between the two quantons. The momentum operator is $\hat{p} = -i\hbar\nabla$, where ∇ is the gradient.

The state function (or vector) ψ is a complex valued function of the space and time coordinates. Its precise form is determined, up to constants, by the precise form of the hamiltonian $\hat{H}$ as well as by the initial and boundary conditions. In the simplest case, that of a single quanton free from external forces, ψ looks like a classical plane wave; in the case of an attractive central force (such as that exerted by an atomic nucleus on an electron), ψ looks like a classical spherical wave. These purely *formal* analogies gave rise to the misnomers 'wave function' and 'wave mechanics', just as the use of the hamiltonian formalism associated with the Schrödinger equation suggested the misnomer 'particles' for the ultimate specific referents of quantum mechanics. Bohr (1934, 1936, 1949) held that the undulatory and the corpuscular views were mutually complementary, whence we had to keep them both and play dialectical games with them. On the other hand Heisenberg (1930 p. 10) admitted that they are only "mental pictures" and "are both incomplete and have only the validity of analogies which are accurate only in limiting cases". Since they are indeed just analogies, and since they cannot be both correct, we shall adopt neither of them. We hold instead that the central referents of quantum theory are *sui generis* entities deserving a name of their own: *quantons*.

The independent variables and constants occurring in the Schrödinger equation for a single quanton, such as an electron in an electromagnetic field, are as follows. (Taken from Bunge 1977b.)

Symbol	*Concept*	*Referent*
x	Point in space	Space
t	Instant of time	Time
m	Quanton mass	Quanton
e	Quanton charge	Quanton
$\hat{p}$	Quanton momentum	Quanton
$\langle A_0, A\rangle$	Field four-potential	Field
$\hbar$	Planck's constant	-----

Since space and time are not physical things, they do not qualify as physical referents, or at least not on the same footing as the rest. (On a relational theory of space and time they are constructs referring ultimately to changing things: Vol. 3, Ch. 6.) The genuine physical referents of the

Schrödinger equation are then quantons and their environment. Of course the latter may consist of a measuring apparatus interacting with the quanton(s). But this is a special case; and quantum mechanics, being a fundamental theory, is not restricted to special cases. Besides, even if we specialize the hamiltonian operator to represent a quanton interacting with an apparatus, we still deal with a purely physical system. No amount of trickery can succeed in smuggling the concept of an observer into the above list.

A second axiom of quantum mechanics is the *eigenvalues equation* for an operator $\hat{A}$ representing an arbitrary dynamical variable ("observable") A. It reads

$$\hat{A}u_k = a_k u_k, \tag{3}$$

where a_k is the k-th eigenvalue (admissible value) of $\hat{A}$, u_k the corresponding eigenfunction (admissible solution), and k a real number. (We are not requiring that $\hat{A}$ be hermitian, so a_k need not be real. But we are assuming that $\hat{A}$ has a non-degenerate spectrum. This assumption is made for the sake of simplicity and it does not mutilate the philosophical problematics.) The simplest case is that of the linear momentum operator $\hat{p} = -i\hbar\nabla$. Its eigenfunctions are $\exp(ikx)$, where k is an ordered triple of reals; the corresponding eigenvalues are $p = \hbar k$. Another example is $\hat{H}u_k = E_k u_k$, which corresponds to a quanton in a stationary state – e.g. an atom in its ground state. In the case of the hydrogen atom the energy eigenvalues are $E_k = -g/n^2$, where n is a natural number and g the energy of the ground state of the system ($n = 1$).

According to the Copenhagen school, every a_k in (3) is one of the values that an *observer* may find when *measuring* the property A with a suitable instrument – of *any* kind. However, it is plain that (3) makes no allusion whatever to any observers, instruments, measurement techniques, or measurement operations. The only interpretation Equation (3) tolerates is a strict or literal one, namely that *a_k is one of the possible values of A* – whether or not we happen to measure it. So much for the semantic aspect of the question. The methodological aspect is just as clear: the results of a precision measurement depend not only on the thing measured but also on the measurement method, and they are rarely exact. (They can be accurate only if the eigenvalues are denumerable and widely separated.)

In general, a precision measurement of a property A will yield a whole set of values. One usually compresses this multiplicity into a formula such as

$$\textit{meas}\ A = a'_k \pm \varepsilon_k, \tag{4}$$

where a'_k is the arithmetic mean of a set of measured values, and ε_k the corresponding relative error, which is characteristic of the measurement method as well as of k and of the size of that set. As a rule a'_k differs from the theoretical value a_k. If both were always identical, as implied by the Copenhagen interpretation, it would be possible to scrap all the research projects devoted to measuring the eigenvalues of all dynamical variables, since these eigenvalues would be given accurately and a priori, hence once and for all, by Equation (3). Fortunately, the workers at Cern, Dubna and Fermilab need not worry: the eigenvalues equation warrants only an objectivistic interpretation since, by hypothesis, the property A represented by the operator $\hat{A}$ is a property of an individual quanton, not of a microsystem including experimental equipment and experimenters.

We are now ready for our third postulate, the *superposition principle*, which we formulate in line with our objectivistic version of the first two principles. Let a be a quanton in state ψ_a, and let $\{u_k | k \in \mathbb{N}\}$ be the family of eigenfunctions of an arbitrary operator $\hat{A}$ representing a dynamical property A of a. Then every linear combination ψ_a of all the eigenstates of $\hat{A}$, i.e.

$$\psi_a = \sum_k c_k u_k,$$

where $|c_k|^2$ is the weight of the contribution of u_k to ψ_a, represents a possible state of a. Graphically: the c_k are the projections of ψ_a onto the axes u_k of the infinite-dimensional Hilbert space associated with $\hat{A}$.

(The state of a quanton of a given kind is representable as a point, or rather a ray, in a state space. The latter is an infinite-dimensional functional space – a Hilbert space. Just as we can coordinatize ordinary space in infinitely many ways, so every Hilbert space can be associated with a number of special grids. These grids are so many functional spaces, every

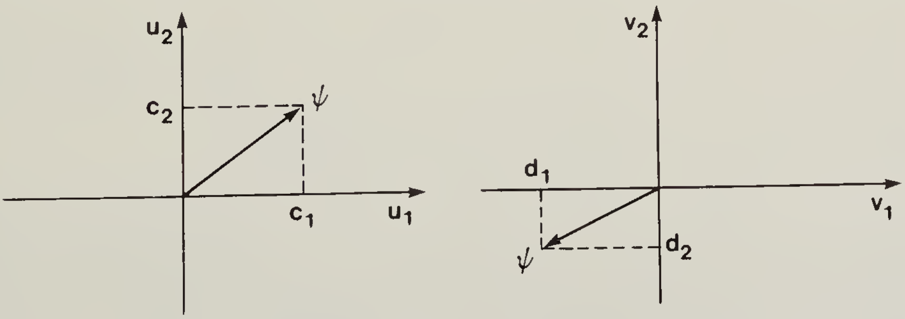


Fig. 2.5. Two different representations of the state ψ of a quanton: $\psi = \sum c_i u_i$ and $\psi = \sum d_j v_j$. The projections are the scalar products $c_i = (u_i, \psi)$ and $d_j = (v_j, \psi)$. Each auxiliary grid has infinitely many axes; in some cases this infinity is non-denumerable.

axis of which is a member of a definite family of functions, e.g. hypergeometric functions, Hermite polynomials, or plain trigonometric functions. And, just as a vector in ordinary space can be decomposed variously, according to the coordinate system that is chosen, so every direction in a Hilbert space can be represented variously in different functional spaces: see Figure 2.5. In other words, one and the same state vector can be projected onto the axes of as many auxiliary grids as desired. In particular, if the auxiliary grid is formed by the axes u_i, the state ψ will be decomposable in this way: $\psi = \sum c_i u_i$, where c_i is the projection of ψ onto the i-th axis u_i. By choosing a different auxiliary grid formed by the axes v_j, different projections are obtained: $\psi = \sum d_j v_j$. The choice of auxiliary grid is conventional, but not all auxiliary grids are mere mathematical fictions. In fact some of them correspond to dynamical variables. More precisely, every dynamical variable can be associated a peculiar functional space, the axes of which correspond to the precise, "sharp", or "well defined" values of the variable. Actually all the dynamical variables that can have simultaneous sharp values can be associated to one and the same auxiliary grid. In short, given quantons of a certain kind, quantum theory assigns them one basic state function together with a set of auxiliary functional spaces, or grids, every one of which is associated with one or more dynamical variables. These auxiliary grids are usually called 'representations'.)

The most popular interpretation of (5) is that it is a purely analytic tool allowing us to cope with our uncertainty concerning the state the quanton really is in. Actually, it is argued, the quanton is always in some definite eigenstate or other, so that its property A has a definite value – only we do not know which. According to this classicist interpretation, $|c_k|^2$ is the subjective probability that a be *found* to be in the eigenstate u_k. This is not our view: we agree with Bohr (1934) that the superposition principle, far from being a mere mathematical trick, has a deep physical meaning, and must therefore be taken seriously: whereas in CP it is often optional, in QP it is mandatory. But we interpret it in strictly physical (objective) terms, i.e. without reference to anyone's ignorance or uncertainty.

As is clear from our formulation, the superposition principle states that in general, i.e. except when the quanton happens to be in an eigenstate of some dynamical variable of it (i.e. when $\psi_a = u_k$), a quanton is in a linear combination of eigenstates. Therefore *in general physical properties lack sharp values*. Instead, each property A has an entire range of values, and each value a_k of A occurs in that range with weight $|c_k|^2$. In other words, generally $|c_k|$ is not sharply peaked. When it is sharply peaked – an exceptional case

– the situation corresponds to the classical case, in which the corresponding property has a sharp value.

Example 1 In general quantons are spatially diffuse. The so-called position indeterminacy or uncertainty Δx, which is a measure of the objective fuzziness in position, is rarely zero. (More in Sect. 5.1.) Since generally quantons are not at precise points in ordinary space, they do not have precise orbits either. Thus, according to quantum theory shape and bulk are not basic universal properties. This is why they play only a minor role in our ontology (Vol. 3, Ch. 6, Sect. 2.4). *Example 2* In general an electron is in neither of its two spin eigenstates $\frac{1}{2}\hbar$ (up) and $-\frac{1}{2}\hbar$ (down), but in a linear superposition of these states. (Classical counterpart: the precession movement of a spinning top in a gravitational field.) If the electron happens to go through an inhomogeneous magnetic field, either in nature or in the laboratory, its unpolarized state is projected onto either of the two spin states. More in Sect. 6.1. *Example 3* Classically, an unstable system is in either of two states: undecayed or decayed. Quantum-theoretically, as soon as such a system (e.g. a neutron) is formed, and as long as it has not disintegrated, it is in a superposition of the decayed states. More on this in a moment.

We take the superposition of eigenstates to be *typical of quantons* in contradistinction to classons. In other words, we regard quantons as being characteristically and objectively fuzzy. This point is crucial in order to understand one of Einstein's objections to quantum theory as well as the recent claim that experiment, by confirming the theory, has refuted scientific realism. Let us deal quickly with the former, leaving the latter for Sect. 6.2.

In their famous criticism of quantum mechanics, Einstein, Podolsky and Rosen (1935) took it for granted that every property of a physical object has a sharp value at any given instant, whether or not we know it. Bohr (1935, 1949) objected that it makes no physical sense to speak of things and their properties in themselves, i.e. apart from observation acts, for to be is to be measured. (For a criticism of this semi-Berkeleyan epistemology see Bunge, 1955a, 1959b.) In our view Einstein and his coworkers were right in asserting the realistic thesis that the world exists without our assistance. But they were wrong in supposing that the world is composed exclusively of things all the properties of which have sharp values (i.e. eigenvalues) at all times. This hypothesis is not *realist* but *classicist*: it amounts to claiming that the world is composed of classons, and therefore must be describable by classical (or neoclassical) theories.

We must distinguish realism from classicism. Our own view is *realist but* *not classicist.* In fact we accept quantum mechanics but not its semi-subjectivist interpretations, in particular that of the Copenhagen school. We admit then that physical properties are of two kinds: definite or sharp, and indefinite or fuzzy. Thus, according to non-relativistic quantum mechanics (though not according to alternative theories), mass and charge are sharp in all circumstances. On the other hand position, momentum, angular momentum and energy are normally indefinite or fuzzy, in the sense that their values are normally ranges of numbers (or of n-tuples of numbers) rather than single numbers. This fuzziness has nothing to do with our ignorance or with the perturbations caused by some observation instruments. Observation merely confirms the *inherent fuzziness* of quantons. And experiment may reduce fuzziness in some respect (e.g. localization) while increasing it in another (e.g. velocity): see Sect. 5.1.

Schrödinger (1935a) thought he had refuted the superposition principle with an ingenious paradox. A closed box contains a cat and a cat-killing device activated by the disintegration of a radioactive atom. The latter has a decay probability of $\frac{1}{2}$ within the hour; hence the probability that the cat be dead after one hour is $\frac{1}{2}$. Schrödinger reasoned that, according to the Copenhagen school, as long as no one opened the box, the cat should be in a superposition of its living and dead states. This superposition would collapse on to either the living or the dead state, with equal probability, on opening the box: the observer creates the event.

Schrödinger claimed that such a state function cannot represent the objective states of affairs but only our imperfect knowledge of it since, whether we know it or not, the cat is either alive or dead. Ergo the corresponding probabilities are actually subjective not objective. And the superposition principle would be merely a mathematical trick without a real counterpart. (Moreover, all the properties would have sharp values at all times even while not being measured.)

The experimental physicist is likely to see no paradox, for he can install an automatically operating film camera inside the box and develop it afterwards to watch the course of events. Nor is the theoretical physicist likely to be persuaded by Schrödinger, for he may argue that the "live" and "dead" states are states of a *macro*system (the cat-in-the-box) that fails to satisfy the Schrödinger equation – so much so that the very concepts of life and death make no sense in quantum theory. (Even if one were to set up and solve the Schrödinger equation for that system, obtaining the solutions ψ_L and ψ_D, and the superposition

$$\psi = 2^{-1/2}(\psi_L + \psi_D), \tag{6}$$

one could reason as follows. At any given time, if ψ_L is non-zero, ψ_D vanishes, and conversely, so that the interference term $\psi_L^*\psi_D + \psi_L\psi_D^*$ vanishes at all times, which confirms that we have to do with a classical system not with a quanton. Another way of realizing that this is so is to sketch a subspace of the state space of the system. One then realizes that the *L* and *D* states are widely separated, so they cannot interfere with one another: see Figure 2.6. Furthermore, since the entropy of the dead cat is so much greater than that of the living cat, the state function of the dead cat – if it could be written out at all – would be a mixture rather than a pure state like (6).) In sum *there is no cat paradox*, but just an illegitimate application of the superposition principle.

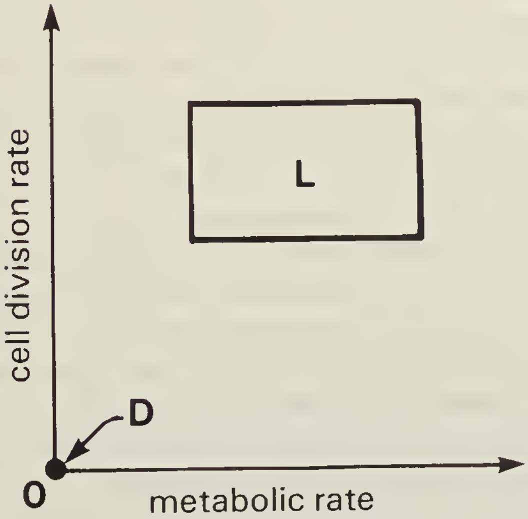


Fig. 2.6. Schrödinger's cat. Subspace of cat's state space formed by two typical biological functions. Point *D* (origin) represents dead state, region *L* represents all possible living states. *D* and *L* cannot interfere. Cats are classons not quantons: the superposition principles does not hold for them.

So much for the first three postulates of quantum mechanics. The fourth will be examined in the next section.

5. CHANCE

5.1. *Probability*

So far we have interpreted ψ as the state function of a quanton, whether simple or complex. Knowledge of ψ summarizes our knowledge of the quanton concerned, as least so far as quantum theory is concerned. However, the representation of a state afforded by ψ is roundabout. Like lagrangians, hamiltonians and partition functions, ψ is a source function out of which particular properties can be constructed. One of these is the position probability density. That this is indeed so is what our fourth and last postulate states.

Born's postulate assigns a physical meaning to the state function ψ. It is a *semantic* postulate, not a law statement; what is a law statement is Schrödinger's equation (1) *together* with Born's axiom. We formulate Born's axiom as follows: "Let ψ_a be a solution of the Schrödinger equation for a quanton a. Then the probability that a be at time t in the region of space comprised between x and $x + \Delta x$ equals $|\psi_a(x, t)|^2 \cdot \Delta x$." The probability in question is a property of thing a in its environment, and therefore it is an objective probability. (For the objectivist concept of probability see Ch. 1, Sect. 4.2.) More precisely, it is the strength of the tendency or propensity that, when in state ψ_a, a be present in the volume element Δx situated at the tip of the vector x. Here again there is no suggestion of measurement: the principle is so general that it is stated for arbitrary quantons with whatever hamiltonian operators they may be assigned.

The Copenhagen school has its own version of the principle, namely: "$|\psi_a(a, t)|^2 \cdot \Delta x$ is the probability of *finding* a in Δx when the position of a is *measured* with the help of an *arbitrary* position-measurement device". This interpretation is illegitimate because ψ_a need not contain any measuring device coordinates – unless of course the corresponding hamiltonian $\hat{H}$ happens to contain them. But since the principle is utterly general, we should not care what $\hat{H}$ may look like, hence what the corresponding ψ_a may be. Moreover, any experimental physicist knows that the probability of finding a thing at a given place is not an intrinsic property of the thing: it depends not only on the probability of the thing being there, but also on the sensitivity of the search instrument and the skill of the experimenter. Thus, the probability of finding a needle in a haystack depends crucially on whether or not the explorer is wearing his eyeglasses and uses a metal detector. In summary, the operationist version of Born's semantic postulate is mathematically illegitimate because it assumes coordinates that are not there; and it is methodologically false because it construes "observable" as a unary predicate.

According to the standard interpretation, ψ is "merely a subjective computational tool" (Kemble 1937 p. 328); moreover, ψ serves only to predict the possible outcomes of experiments done simultaneously or successively on similarly prepared quantons. On the other hand, in our realistic interpretation Born's axiom involves *objective* probabilities that may refer to *single* quantons and, moreover, to quantons not subjected to any experimental scrutiny. That the probabilities are objective is clear from the arguments of ψ, none of which refers to a knowing subject. (Remember from Sect. 4.2. that the Schrödinger equation, which ψ solves, contains no information about any observers.) To be sure the computed probabilities may turn out to be false, so that we may have to alter our premises in the light of new empirical information. But in this regard working in quantum theory is no different from working in classical physics. (See Newton 1980 and Ch. 1, Sect. 4.2.)

That the probability associated with the state function may refer to a single quanton, rather than to a statistical ensemble, is obvious for a similar reason: unless ψ does solve the Schrödinger equation for an ensemble of coexisting quantons, such as those composing a metallic body or a laser beam, *by hypothesis* it refers to a single (though possibly complex) quanton a, such as a molecule. This point is often misunderstood because the *test* of any calculated probabilities for single cases involves some ensemble of similar (coexisting or successive) entities. However, we must not confuse reference with test even though we must relate them. (See Vol. 1, Ch. 2, Sect. 2.6., and this volume, Ch. 1, Sect. 4.2.)

When a theoretical (e.g. probabilistic) statement is subjected to empirical tests, it must first be operationalized: i.e. it must be conjoined with indicator hypotheses and translated into laboratory or field jargon (Vol. 6, Ch. 11, Sect. 2.2 or 1973a, Ch. 10). Thus a quantum theoretical statement such as "The probability that individual thing a, of kind K, be in state s, is p", will have to be translated into something like this: "If an observation (or measurement) is performed on an ensemble (collection) of entities of kind K prepared in a similar fashion, a fraction near p of them should be found (if the theory is true) to be in state s".

If the difference between this statement, which refers to a collection under experimental control, and the previous statement, which refers to an individual, is overlooked, then one will mistakenly attribute quantum theory a *statistical* character rather than a *probabilistic* one: i.e. he will infer that quantum theory is always about collections (in particular Gibbs ensembles) of similar entities, never about individual things. (This is in fact the statistical interpretation of quantum mechanics originally proposed by Slater (1929) and favored by Einstein (1948), Blokhinzev (1964), Ballentine (1970), de la Peña (1979), and a few others.) As we saw above, this opinion on the referents of quantum theory contradicts the fact that in many cases ψ is made to refer, *by hypothesis*, to a single quanton in a given environment – e.g. an electron in an electromagnetic field, or even in free space. It also contradicts the very existence of quantum statistical mechanics, which does in fact refer exclusively to ensembles of coexisting quantons. And it is *experimentally refuted* by electron diffraction experiments, which can be done with a single electron at at time (Sect. 5.2). For all these reasons we should speak of the *probabilistic*, not the *statistical*, interpretation of quantum theory mediated by Born's axiom.

(Some people distrust the very notion of the probability of a thing being in a given state. In some cases the objection stems from the classical principle that every thing in in some sharp state, i.e. an eigenstate, at all times, so that the very notion of a linear combination of states makes no sense to them. At other times the objection originates in the operationist thesis that we can measure only frequencies of changes of state, such as the emission or absorption of a photon by an atom. The unease should disappear upon reflecting that (*a*) quantum theory calculates the probabilities of states, not only of events, and (*b*) the probability of a thing being in a given state can be construed, and sometimes even calculated, as the sum of the probabilities of all the possible processes, or paths, leading to that state, i.e. as the sum of all the transition probabilities.)

Our objectivist version of Born's principle clarifies the following point on position coordinates. Classical mechanics contains a position coordinate X that assigns to each particle, relative to each reference frame, at each instant, and on each distance unit, a triple of real numbers, namely the components of the position vector along three coordinate axes. (I.e. X: $P \times F \times T \times U_d \rightarrow \mathbb{R}^3$. Similarly for points in a field.) Quantum mechanics contains no such function. The position operator $\hat{x}$ is not a vector the components of which have definite values; only its probability density $|\psi|^2\hat{x}$ has sharp values at all times. Hence quantum theory *cannot assign any precise positions* to quantons. (Mathematically: $\hat{x}$ has no well-behaved eigenfunctions, hence no well defined eigenvalues.) And, since quantons have in general no sharp positions, they have in general *no sharp trajectories* either. (To be sure one can compute $d\hat{x}/dt$ to obtain $[\hat{p}-(e/c)A]/m$. However, this entails only that the eigenvalues and the mean values of the velocity are those of the corresponding kinetic momentum divided by the mass. We cannot integrate the preceding equation of motion to obtain the quanton trajectory. Again, since there is no sharp position there is no sharp orbit.) In brief, quantons are not point-like. Nor are they wave-like. (This can be seen e.g. by comparing the Galilei transform of a classical wave with that of a ψ "wave". Whereas the former remains unchanged, the latter does not. Indeed, the transformed ψ "wave" contains terms representing the momentum and kinetic energy of the moving frame.)

The probability density $|\psi(x, t)|^2$ allows one to calculate all the statistical parameters defined in probability theory, in particular averages, momenta, and mean standard deviations. In particular we can compute the mean standard deviation of the distribution of values of any dynamical variable, such as the position and the momentum of a quanton in a given state. (The average of a dynamical variable A is $\langle \hat{A} \rangle = (\psi, \hat{A}\psi)$, the nth momentum of A is $\langle \hat{A}^n \rangle = (\psi, \hat{A}^n\psi)$, and the mean standard deviation of $\hat{A}$ is given by $(\Delta\hat{A})^2 = \langle \hat{A}^2 \rangle - \langle \hat{A} \rangle^2$.) Moreover, if we settle for inequalities we can obtain certain results of the utmost generality, i.e. without making any precise assumptions about the nature of the quanton(s) concerned, except that they "obey" quantum theory. Let us do this in the case of the most celebrated and misinterpreted of all the formulas of quantum theory.

Heisenberg's inequalities relate the mean standard deviation $\Delta\hat{x}$ of the position, and $\Delta\hat{p}$ of the momentum, of an arbitrary quanton in an arbitrary state at any given time. They read:

$$\Delta\hat{x} \cdot \Delta\hat{p} \geq \hbar/2 \tag{7}$$

for every component. The preceding formula is a rigorous consequence of only three general premises: (*a*) the commutation equation $\ulcorner \hat{p}\hat{x} - \hat{x}\hat{p} = -i\hbar \urcorner$, obviously satisfied by $\hat{p} = -i\hbar\nabla$; (*b*) the definition, recalled a moment ago, of the mean standard deviation of an arbitrary dynamical variable, and (*c*) a certain mathematical inequality (Schwartz's) devoid of physical meaning.

To derive (7) we need not know anything about the quanton except that it "obeys" quantum mechanics. (We do need a precise knowledge of the quanton and its environment if we wish to obtain an equality.) In particular, no measurement apparatus, much less an observer, is referred to in the premises leading to (7). The observer and the apparatus occur only in popular heuristic presentations as well as in their philosophical discussions.

All (7) "says" is that *the width $\Delta\hat{x}$ of the position distribution is inversely proportional to the width $\Delta\hat{p}$ of the momentum distribution*, and this regardless of the type of quanton and the measurement procedure: see Figure 2.7. Whatever else one may wish to read into (7) is sheer fabrication: it has not a mathematical leg to stand on and can only be justified by philosophical bias.

In the Copenhagen camp there is no consensus on the way (7) is to be interpreted – an indicator of the uncertainty that characterizes that interpretation. The most popular interpretations of (7) are the following five:

(i) The formula represents *mutually complementary measurement devices*: anything that measures x with any precision destroys the very possibility of an exact measurement of p and conversely. *False*, because the premises that

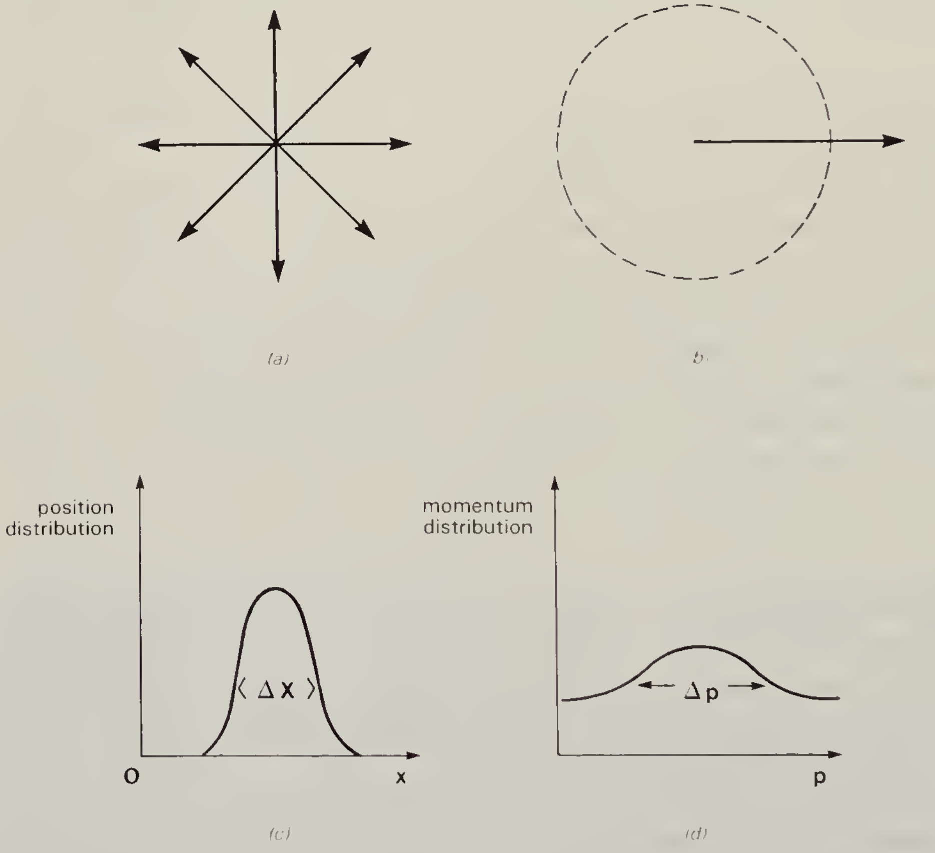


Fig. 2.7. Heisenberg's ("indeterminacy", "uncertainty") inequalities ("relations"). (*a*) A fairly well localized quanton moves at all speeds in all directions. (*b*) A quanton moving at a definite speed is poorly localized (unless constrained by walls or fields). (*c*), (*d*) The smaller the position spread, the greater the momentum spread.

entail (7) make no reference whatsoever to any measurements: they hold for all quantons in any states, whether free or acted upon by measuring instruments.

(ii) The mean standard deviations related by (7) are *caused* by the measurement act, for which the smaller the object the greater the disturbance. *False* for the above reason and also because it presupposes the classical view that the properties of all physical things have sharp values at all times.

(iii) The deviations occurring in (7) are plain measurement errors. *False* for the two reasons given when examining (ii).

(iv) The dispersions involved in (7) are *widths of the wave packet* (in ordinary space and in momentum space respectively). *False* because ψ does not represent a wave. (Electrical engineers are familiar with an analogous inequality that holds for classical waves, but this is another story.)

(v) The standard deviations are *uncertainties* in the physicist's mind concerning the exact position and momentum of the quanton. *False* because (7) refers to quantons, not to human brains, and because the assumption that quantons have precise positions and momenta at all times (only we do not know them exactly) is alien to quantum theory. Uncertainty is a feature of the Copenhagen school, not of quantons.

In short, all five popular interpretations of Heisenberg's theorem are unwarranted by the premises that entail it. (Moral: Thou shalt not talk in the theorems whereof you kept silent in the axioms and definitions. More bluntly: Don't smuggle.) Heisenberg's inequalities represent an objective non-classical property of quantons that has nothing to do with measurements or mental states. Being essentially probabilistic, they refute classical (Laplacian) determinism in both its ontological and its epistemological versions. (According to the former the future is uniquely *determined* by the present; according to the latter the future can be *predicted* exactly from an accurate knowledge of the present.) However, (7) does not justify radical indeterminism or the denial of lawfulness, because (7) happens to be a law statement. (See Bunge 1959a for general determinism.)

Heisenberg's indefiniteness inequalities show, among other things, that quantons are not point particles. (Another argument is that the eigenfunction of the position operator is a delta, which has an infinite norm and is therefore to be excluded from the formalism.) Not being point particles, quantons have no precise trajectories. But they do have average trajectories that coincide with the classical ones. (Roughly, a classical orbit is the trajectory of the center of the probability distribution.)

The Heisenberg inequality (7) is a special case of a general formula. Let $\hat{A}$ and $\hat{B}$ be dynamical variables (in either quantum mechanics or quantum field theory) satisfying a commutation equation of the form

$$\hat{A}\hat{B} - \hat{B}\hat{A} = i\hat{C} \tag{8}$$

where $\hat{C}$ is a hermitian operator, i.e. one with real eigenvalues. (Examples: the components of the orbital angular momentum and of the spin.) By a reasoning similar to the one resulting in (6) one obtains

$$\Delta\hat{A} \cdot \Delta\hat{B} \geq \tfrac{1}{2}|\langle \hat{C} \rangle| \tag{9}$$

where $\langle \hat{C} \rangle$ is the (spatial) average of $\hat{C}$. Pairs of dynamical variables that satisfy equations of the form (8) are said to be *conjugate* to one another. They are also said to be *mutually complementary*, in the sense that, the sharper or more peaked the distribution of one of them is, the more widely distributed its conjugate is.

The usual operationist or Copenhagen interpretation of the commutation equation (8) and its consequence (9) is this: "If two 'observables' (dynamical variables) are not simultaneously measurable, then their corresponding operators do not commute, and conversely. In other words, non-commutativity amounts to lack of simultaneous measurability". We reject this interpretation because, as noted above, no reference to measurement is made in either (8) or (9). Our own interpretation of Heisenberg's generalized commutation equation (8) and inequality (9) is this: "If two dynamical properties of a quanton fail to *have* sharp or definite values at the same time, then their corresponding operators do not commute. (Put it negatively: a quanton is never in the eigenstates of non-commuting dynamical variables.) And, not *having* sharp simultaneous values, no measurement can *find* them: we cannot measure something that is not there".

Our realist interpretation of (8) and (9) turns the Copenhagen interpretation upside down: *it is not that measurement precludes our finding simultaneous sharp values, but that the non-existence of the latter prevents us from measuring them.* Our interpretation is also at variance with classical realism (of the Einstein-de Broglie-Bohm type), according to which quantons do have sharp properties (represented by hidden variables) at all times, only quantum theory is incomplete, so a new theory, including hidden variables, is needed. In our view this is only wishful thinking and a hang-up from classical physics. (We shall return to it in Sect. 6.2.) The basic constituents of the universe are somewhat fuzzy, yet just as real as the neat macrosystems they constitute.

The Heisenberg inequalities (7) constitute the most popular example of the general formula (9). Another important though less well known case occurs in quantum field theory: it involves the operator $\hat{N}$ representing the number of field quanta (e.g. photons) and the operator Θ representing the corresponding field phase. These two are conjugate dynamical variables: they are related by $\hat{\Theta}\hat{N} - \hat{N}\hat{\Theta} = i$. (More precisely, $\hat{\Theta}_{k\lambda} \cdot \hat{N}_{k'\lambda'} - \hat{N}_{k'\lambda'} \cdot \Theta_{k\lambda} = i\delta_{kk'} \cdot \delta_{\lambda\lambda'}$, where k and λ are the wave and the polarization vectors respectively.) The corresponding standard dispersions inequality is, in accordance with Equation (9),

$$\Delta\hat{\Theta} \cdot \Delta\hat{N} \geq \tfrac{1}{2}. \qquad (10)$$

In words: the scatters in the particle number and in the wave phase are mutually dual or complementary. The ideal limit cases are

$\Delta\hat{N} \to 0, \quad \Delta\hat{\Theta} \to \infty$	a sharp or well defined particle number but no definite wave;
$\Delta\hat{\Theta} \to 0, \quad \Delta\hat{N} \to \infty$	a sharp or well defined wave but no definite particle number.

These two are only ideal cases: actually quantons behave neither as particles nor as waves. If preferred, every quanton is in a superposition of particle and wave states.

We take Equation (10) to be a precise formulation of the *complementarity principle* (actually a theorem), stating that the particle and wave *pictures* of matter are mutually complementary (Bohr, 1934, 1936, 1949). Actually Bohr did not state the "principle" in relation to (10) but to (7), and he did not say that it shows the dual nature of *matter* but rather that it exhibits the complementarity between different types of *experimental set-ups* (Bohr 1949). However, (7) does not contain any undulatory property. And (10) cannot be interpreted in terms of measurements, if only because phases are not measurable: only phase differences are. (Interestingly, the phase difference $\hat{\Theta}_1 - \hat{\Theta}_2$ between the two quantons and their total particle number $\hat{N}_1 + \hat{N}_2$ commute. Hence we can measure accurately and at the same time their phase difference and the total particle number, but we cannot know which "particles" go with which "waves".)

Bohr's complementarity "principle" (10) has been endlessly discussed in the literature for being regarded as the clue to the interpretation of quantum theory. However, there is no consensus on either its formulation or its significance. Indeed the following alternative versions of it have been defended: (*a*) it describes the dual, or even contradictory, nature of *matter*, i.e. the fact that on the quantum level things are both corpuscular and undulatory; (*b*) it expresses the complementarity of *measurement set-ups*, i.e. the fact that, whereas certain devices measure corpuscular properties, others measure undulatory ones; (*c*) it expresses the duality and complementarity of the particle and wave *pictures* of nature. (That the three alternative views are different and are usually conflated is hardly noticed.)

In our view each of the three versions of the complementarity "principle" contains a grain of truth and none is central to quantum theory because, as shown by (10), the particle and wave "aspects" (of quantons, measurement devices, and descriptions) are only *ideal limits* and moreover *mutually exclusive ones*. (This destroys, in particular, the thesis that complementarity illustrates dialectics.) We have come to understand that quantons are much too rich to be describable in classical terms such as "particle" and "wave". We have learned that the referents of quantum theory are neither corpuscles nor fields, although they do possess, in certain *special* cases, particle-*like* properties, and in others field-*like* properties – e.g. photons transport quanta of energy, and electrons can diffract. So, although we retain the complementarity "principle", we reinterpret it in objective terms and assign it a modest role.

The following analogy, (Lévy-Leblond and Balibar 1984 p. 67), illuminates the situation. A cylinder presents both circular and rectangular features, but it is neither a circle nor a rectangle. And saying that the circle and the reactangle are mutually complementary will not help understand what a cylinder is – whereas exhibiting a solid cylinder or the corresponding mathematical formula will help. Yet it is true that, in exceptional circumstances – namely when seen along its axis – a cylinder looks *as if* it were a circle; and in others – namely when looked at from the side – it *appears* as a rectangle. Likewise a quanton appears as a particle when it collides with (actually is scattered by) another quanton; however, the collision cross section is anything but the classical (geometrical) one, for it depends upon the wavelength of the quantons. And in other circumstances, e.g. when diffracted by a crystal lattice, a quanton *appears* as a field; however, a calculation of the diffraction pattern involves knowledge of its mass and spin if any. The truth about quantons is told by quantum theory (in its realistic interpretation), not by partial classical analogies that were inevitable, and even had some heuristic value, in the early confused days of the theory.

Finally we must say a word about the popular view that the Heisenberg inequalities and their kin set a *limit on our knowledge* of nature by showing, e.g., that we cannot know at the same time the exact position and the exact momentum of an electron. Obviously, this interpretation presupposes that the electron *has* an exact position and an exact momentum at all times. If this were true then our knowledge would indeed be deficient and incapable of being improved on. However, that is not what quantum theory states: as we saw a while ago, quantons have distributions of position and momentum (and angular momentum and energy, etc.): sharp values are exceptional and moreover ideal cases. Hence *there is no sharp value to be known*: we cannot undertake to find out what is not there. In other words, the popular view we are discussing presupposes that the referents of quantum theory are classons, not quantons. And it overlooks the fact that Heisenberg's inequalities and their kin do not refer to our knowledge but are (derivative) law statements referring to physical objects whether or not they are under observation. In sum, Heisenberg's inequalities and their kin, far from setting limits upon human knowledge, constitute an important addition to it.

5.2. *Double Slit and Double Logic*

Let us now investigate a puzzling problem that is usually treated in an either a classical or a subjectivistic manner, and that has occasionally suggested the need for a change in logic: the *double slit experiment*. (See Figures 2.8

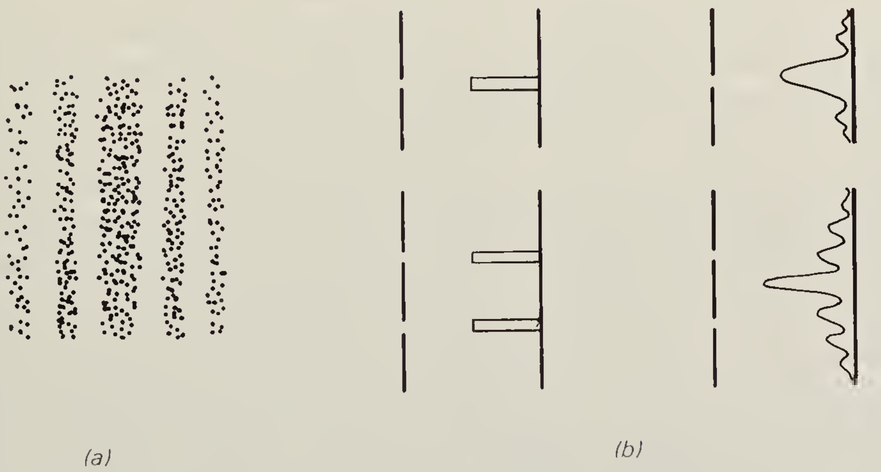


Fig. 2.8. "Interference" patterns. Quantons are emitted by a source, pass through a slit system, and strike a detecting screen, e.g. a photographic plate. (*a*) The "interference" pattern of individual impacts of quantons passing through a double slit system. (*b*) Comparison of behavior of classical corpuscles and classical waves. See Jönsson (1974) for electron diffraction photographs from a single slit and multiple slits. Caution: the intensity curves occurring in that paper were calculated classically using Huyghens's principle.

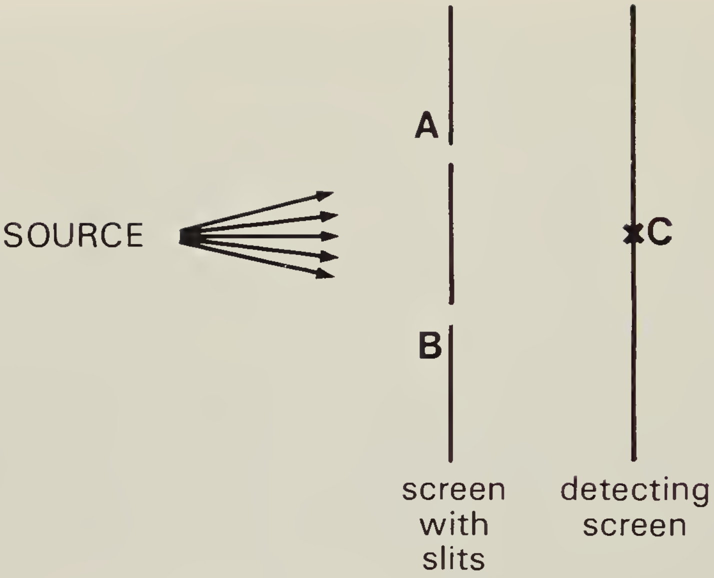


Fig. 2.9. The double slit experiment as alleged empirical refutation of distributive logics and hence empirical support for quantum logics. Source of fallacy: the dogma that quantons are particles.

and 2.9.) If quantons (e.g. electrons) are emitted one at a time (e.g. by an electron "gun"), those that make it through the slit system strike the recording screen producing a flash (fluorescent screen) or a black dot (photographic plate). The distribution of impacts on the recording screen is not the superposition of the distribution resulting from the single slits (i.e. when one of them is closed). A qualitatively new pattern emerges, that is explained in terms of the interference of real waves in the case of light, and the interference of ψ "waves" in the case of electrons and other massive quantons. These distributions are statistical or mass outcomes: the individual thing, be it photon or electron, is recorded as a small (though extended) spot. As the intensity of the incoming beam increases, the individual spots arrange themselves into bands that increasingly resemble the interference bands produced with high intensity light sources. That is, the wave aspect of matter (whether massive or not) is an *emergent property possessed by aggregates of quantons* and absent from the individual quantons. The individual quantons behave neither as classical particles nor as classical fields. (Recall Sects. 4.1 and 5.1.)

The double slit experiment has been interpreted in a subjectivistic way, as proving that it is up to the experimenter to "conjure up" now the particle aspect, now the wave aspect, by varying at will the source intensity or the slit apertures. It is certainly true that the experimenter has this power – but so has Mother Nature, which owns plenty of crystals and natural gratings that do the same tricks without deliberation or research grants. The particle and wave aspects are objective properties of matter. But they are only the two ends of a rich gamut of behavior that cannot be compressed into the classical framework. (Recall Sect. 5.1., in particular formula (10), showing that the more pronounced the corpuscular aspect, the less enhanced the wave aspect and conversely.)

Through which slit does the quanton go? According to the strict Copenhagen interpretation, this question is meaningless unless one makes a position measurement – e.g. closes one of the slits, thus making sure that the quanton has passed through the other. (See e.g. Schiff 1947, Heisenberg 1958.) But of course if we do this the question evaporates; so, it cannot be answered at all. The answer given by other authors is different, namely this: When both slits are open, the quanton goes through both of them in such a manner that there is interference between the two waves emerging from them. (Each quanton is said to interfere only with itself: Dirac 1958 p. 9) We agree that the quanton goes through *both* slits and that, for strong beam intensities or long exposure times, the process can be *mimicked* by classical waves. (In fact the interference patterns are usually calculated using Huyghen's principle for light waves, not by solving the Schrödinger equation, which has no closed form solution in this case.) However, we do not admit that the process is one of classical diffraction, for the state function ψ does not represent a classical wave.

We submit that the correct answer is obtained using the superposition principle: when the two slits are open the state of the incoming quanton is a linear combination of the states corresponding to the two slits. To put it in symbols suggested by Figure 2.9:

$$\psi = a\psi_A + b\psi_B, \quad \text{with} \quad a, b \in \mathbb{C}. \tag{11}$$

The resulting probability density is then

$$|\psi|^2 = |a\psi_A|^2 + |b\psi_B|^2 + a^*b\,\psi_A^*\psi_B + ab^*\,\psi_A\psi_B^*. \tag{12}$$

The first two terms are the quantum correlates of the corresponding classical ones, and the last two are the so-called interference terms. However, this is only a mathematical analysis: quantum theory does not authorize us to interpret the first two terms in a corpuscular manner and the last two in an undulatory one. The physical state is not split. The two components of ψ

in (11) are *inseparable* even though they are distinguishable. (More on inseparability in Sect. 6.2.)

Finally, let us examine what happens if one assumes that quantons are particles, so that every quanton goes either through one of the slits or through the other but never through both. In this case one is tempted to reason directly on probabilities rather than probability amplitudes (state vectors). One starts by identifying the three events of interest as suggested by Figure 2.9:

$A = \ulcorner$The quanton passes through slit $A\urcorner$.
$B = \ulcorner$The quanton passes through slit $B\urcorner$.
$C = \ulcorner$The quanton arrives at some point C on the recording screen$\urcorner$.

The probability of the quanton arriving at C on the recording screen is then

$$P(C) = P(C \,\&\, (A \vee B)). \tag{13}$$

If one assumes, in line with ordinary logic, that conjunction distributes over disjunction, one can expand the RHS of the previous equation, obtaining

$$P(C) = P(C \,\&\, A \vee C \,\&\, B). \tag{14}$$

Now, since the quanton is assumed to be a particle, it will go through either slit A or slit B but not through both. In other words the possibilities $C \,\&\, A$ and $C \,\&\, B$ are *mutually exclusive*, so that (14) becomes

$$P(C) = P(C \,\&\, A) + P(C \,\&\, B). \tag{15}$$

However, this formula is false, because it does not contain the symbols representing interference: recall Equation (12). Putnam (1976), Friedman and Putnam (1978) and others conclude that the root of the trouble is the hypothesis of *distributivity* employed in deriving (14) from (13). But physicists never bother with logic; moreover they employ classical mathematics, which has ordinary (distributive) logic built into it. On the other hand physicists calculate probabilities indirectly, namely from state vectors, not directly. And they do not assume that the possibilities $C \,\&\, A$ and $C \,\&\, B$ are mutually exclusive, so they do not obtain the false formula (15). In short, the claim that the double slit experiment forces one to give up distributivity, and hence to adopt a quantum logic, presupposes that quantons are classical particles. (See Selleri and Tarozzi 1978.) We thus confirm a conclusion arrived at in Ch. 1, Sect. 3.2 through a different route: *the logic underlying quantum mechanics is classical.* Therefore quantum logics are idle. (The same conclusion is obtained by axiomatizing quantum mechanics and showing that it employs only classical mathematics. See Bunge 1967c.) What are nonclassical are the quantons, and this is why we need quantum physics rather than classical physics to account for them.

The upshot of this section is as follows.

(i) Quantum physics accounts for real entities – quantons – unknown to classical physics;

(ii) for all its novelties,quantum physics dovetails partially with classical physics;

(iii) quantum theory is substantially correct, but the standard (Copenhagen) interpretation of its mathematical formalism, which is semi-subjectivist, is false and must be replaced with a realist interpretation;

(iv) quantum theory refers to single quantons or systems of such, the states of which it describes by state functions (vectors) that are not directly observable;

(v) in quantum theory probabilities cannot be assumed or calculated from other probabilities but must be calculated ultimately from state functions; the quantum-theoretical probabilities are objective;

(vi) the most general state function of a quanton is a superposition of states corresponding to sharp values of some property of the quanton, such as its energy; sharp values are the exception;

(vii) the state function extends over a region of space that depends on the nature of the environment of the quanton; e.g. an electron in a box occupies the entire box (though it is not homogeneously distributed in it), and an electron diffracted by a two slit screen passes through both slits;

(viii) the state of a quanton is extremely sensitive to environmental variations, whether natural or artificial, which can now enhance the "corpuscular aspect", now the "undulatory aspect" of the quanton;

(ix) the observer is not among the referents of quantum theory; and the apparatus occurs only when it is explicitly represented in the hamiltonian, hence also in the state function, of the system;

(x) the logic underlying quantum theory, whether in its Copenhagen or in its realistic interpretation, is classical logic: quantum logics are only games.

The following section, dealing with two other unclassical ("paradoxical") features of quantons, will reinforce the above conclusions.

6. REALISM AND CLASSICISM

6.1. *Measurement and State Projection*

The main epistemological problem about quantum theory is whether it represents real (autonomously existing) things, and therefore whether it is compatible with epistemological realism. (The latter is the family of epistemologies which assume that (*a*) the world exists independently of the knowing subject, and (*b*) the task of science is to produce maximally true conceptual models of reality. See Vol. 6, Ch. 15, Sect. 2.2, and note that the question of reality has nothing to do with scientific problems such as whether all properties have sharp values and whether all behavior is causal.)

Except for a few unfortunate madmen, people – even quantum theorists and philosophers – go about their daily business taking epistemological realism for granted: consistent non-realists, when found, are locked up and treated. In particular, experimentalists handle pieces of apparatus which they use to observe or transform certain physical objects, without ever doubting that at least those bits of laboratory equipment are real – and with them their microphysical components. (See Born, 1953.) And theorists keep trying to improve the adequacy of our ideas by modifying their theories, rather than engaging in incantations aimed at subduing reality to their preconceived ideas. The very notions of objective test and objective truth presuppose the basic principles of epistemological realism.

Yet one often reads in authoritative books and papers that quantum theory has refuted epistemological realism, at least with regard to microphysical entities. (Those texts never question the reality of the apparatus or the observer, although they admit tacitly that their ultimate components have no autonomous existence.) The main pillar of the non-realist interpretations of quantum theory is a certain view on measurement and on the projection (reduction) of the state function that is involved in measurement. (See Wheeler and Zurek Eds. 1983 for the standard view and some criticisms of it.) We shall show that this view is incompatible with both the theory and the practice of measurement.

Although measurement has become commonplace not only in science and technology but also in indutry and trade, and even in the modern home, some theoretical physicists and philosophers hold quaint views about it. (See Sect. 1.1 of this chapter, and Vol. 6, Ch. 11, Sect. 3.1, for general treatments of measurement.) However, if pressed any experimental physicist will admit that every measurement serves either *to find out or to alter* one or more properties objectively possessed by some real or at least putatively real object. Moreover, every experimentalist will inform us that a microphysical measurement, i.e. one performed on a microphysical entity, be it quanton, semiquanton, or classon, is a process in which the states of the microthing become *correlated* with those of the measuring apparatus, so that observing the latter allows one to infer the former. In other words, the measurement process involves an *amplification* or micro-macro "translation"; moreover, the process is *often irreversible.*

Amplification is the price exacted by our comparatively poor sensory apparatus, and irreversibility the price paid for recording. If we were the size of molecules we might not need amplification. And whenever we let our senses do the recording, they engage in some irreversible process such as a photochemical reaction in the retina. The counting of quantons is a familiar example of microphysical measurement. Thus when an alpha-particle hits a ZnS crystal, it causes the emission of a bunch of photons capable of stimulating the human retina in the dark; and when a charged particle goes through a Geiger-Müller counter, it triggers an electron avalanche causing a click, which in turn can stimulate the human cochlea.

The term 'measurement' is sometimes used to denote the *preparation* of an entity – e.g. a steam jet or a molecular beam – that precedes the measurement proper. (For the difference between measurement and preparation see e.g. Margenau 1978 Ch. 12) The preparation may involve some measurements but these are auxiliary and not always required. For example, the production of a steam jet need involve no measurement: a tea kettle suffices.

At other times 'measurement' is misused to denote *any interaction* of an entity with its environment. Thus it has been said that, in the case of the radioactive nucleus shielded from its macrophysical environment, "the experimental apparatus can be identified with the electrons surrounding the nucleus" (Fonda *et al.* 1978 p. 623). It has also been said that the collision of a proton with other particles in an atomic nucleus "essentially constitutes a measurement" (Horwitz and Katznelson, 1983, p. 1184). These are obvious mistakes, for neither the atomic electrons nor the nuclear protons yield measurement results. The mistakes originate in the Copenhagen article of faith that it is senseless to talk about things (e.g. atomic nuclei) in themselves, i.e. apart from any measurement apparatus. And the mistake involves a logical fallacy: from the truth that every measurement involves an interaction it does not follow that every interaction is a measurement.

However, the worst misconception of measurement is its identification with the subjective experience of *taking cognizance* of the outcome of measurement. This overlooks the fact that an essential point of measurement is to find out how things are independently of our perceptual and conceptual errors. (Thus, if we need to know a distance with some accuracy we measure it instead of just estimating it by eye. And if we wish to know whether a certain flying object is a flying saucer or a plasmoid, we attempt to take some measurements of it.) Besides, the subjectivist misconception ignores that measurements can be fully automated, and their outcomes stored in computer memories or printouts that may be consigned to oblivion. Of course in such case the measurements will be useless, but the point is that they are measurements and physical processes not mental ones. This is why they are accounted for by strictly physical theories. (Only some measurement errors call for psychological considerations.) Whether or not we look at the apparatus once the measurement results are in is irrelevant to the outcome of the measurement process. This is what laboratory experience teaches us and what every realistic theory of measurement states. After all, "Quantum mechanics may have some rather bizarre and peculiar features, but it is not a subjective and mystical creed" (Gottfried 1966 p. 183).

So much for some popular misunderstandings. Let us now note an important distinction seldom made in the quantum theoretical literature on measurement although it was drawn long ago by Pauli (1958 pp. 72–73). This is the distinction between measurements of the first kind, or non-disturbing, or *unobtrusive*, and measurements of the second kind, or disturbing, or *obtrusive*. It is usually stated that all microphysical measurements are obtrusive. But this is not so: thus, measuring the wavelength of a light wave is an unobtrusive operation because it does not alter the value of the property being measured. On the other hand measuring the polarization of a light wave by means of a nicol prism is an obtrusive act: the apparatus does not merely render manifest the state of polarization but produces it. A more familiar example of an obtrusive measurement is that of taking the temperature of a small liquid body by inserting a thermometer in it: this procedure may alter the original liquid temperature because of the heat exchange between the two things.

Ideally, the repetition of an unobtrusive measurement can yield the same result – provided sufficient time has elapsed for the recovery of the measuring instrument. On the other hand the repetition of an obtrusive measurement is bound to yield a somewhat different result. In the case of classons the disturbance introduced by the measuring device can be corrected for by supplying the apparatus with a compensation device; and in principle it can always be calculated theoretically and thus be corrected conceptually. This is not generally true of quantons: in this case the disturbance may neither be compensated for nor calculated.

A much-discussed, though perhaps still not well-understood, example of a microphysical measurement is that of the spin of a massive quanton, such as an ion, with the help of a Stern-Gerlach apparatus. A beam of quantons enters an inhomogeneous magnetic field that splits the beam into two roughly equal parts that impinge upon a recording screen: one of them is composed of quantons spinning in the direction of the external field, the other in the antiparallel direction. According to the usual theory "the Stern-Gerlach experiment measures σ_z [the z-component of the spin] without changing it" (Bohm 1951 p. 594). However, there is no evidence for this interpretation: all we know about the incoming beam is its composition. One assumes explicitly that its individual components are in states that are linear combinations of the spin eigenfunctions ("up" and "down"). The Stern-Gerlach apparatus orients and separates the spinning quantons; therefore it helps perform a typical obtrusive (or second kind) measurement.

Evidently, the very notion of an obtrusive measurement presupposes the *reality* of the object of measurement as well as the possibility of *distinguishing* it clearly from the measuring apparatus. (A can act upon O only if both A and O are real as well as different.) Yet the standard (or Copenhagen) interpretation of quantum theory denies precisely both the independent reality of quantons and their distinguishability from apparatus and observer. Moreover, that interpretation ignores the obtrusive-unobtrusive distinction, and insists that every measurement act disturbs the object in an uncontrollable manner. According to this opinion even the mere observation of radiative or radioactive decay would be caused or prevented by the presence of observers. This is not what the quantum theories of such processes state, for they acknowledge the existence of spontaneous decays. And it is not what experiment indicates either. Thus, Shimony and three of his students (Hall *et al.* 1977) performed an experiment to ascertain whether the presence of observers does in fact influence the emission of gamma-rays by sodium-22 atoms. The result was negative: there is no psychokinesis – hence no Copenhagen spirit acting on matter. Clearly, then, the continual reference of the Copenhagen theorists to measurement is rhetorical.

Worse, the standard interpretation of quantum theory conflates apparatus and observer, and holds that the frontier ("cut") between the observer-apparatus system, on the one hand, and the quanton, on the other, is conventional: that it can be shifted at will by the observer. (See e.g. von Neumann 1932.) Both conflations are at variance with experimental practice. Here it is essential to isolate the observer from the apparatus (e.g. by using servomechanisms) in order to minimize the former's disturbances on the latter. (Recall Sect. 1.1.) And it is equally essential to distinguish what belongs to the object from what is contributed by the apparatus: the very

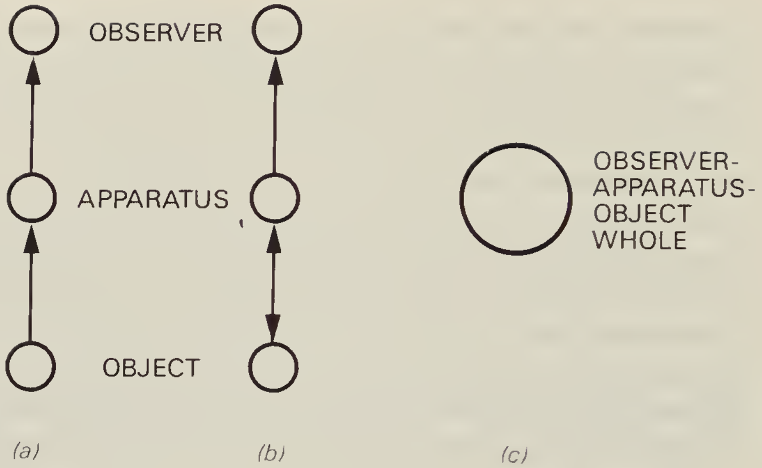


Fig. 2.10. (*a*) Unobtrusive measurement: object send signal to apparatus, which sends signal to observer. (*b*) Obtrusive measurement: apparatus modifies state of object, which sends signal to apparatus, which sends signal to observer. (*c*) Copenhagen interpretation: object, apparatus and observer are indistinguishable except by convention.

notion of an experimental "artifact" (or effect caused by the measuring instrument) rests on the possibility of effecting a sharp distinction between object and apparatus. Neither of the two distinctions is made in the standard quantum theories of measurement (e.g. von Neumann 1932, London and Bauer 1939, Wigner 1963). Both distinctions are made in the realist treatments of measurement (e.g. Everett 1957, Bunge 1967c, Hepp 1972, Machida and Namiki 1980, Cini 1983, Bunge and Kálnay 1983b). The contrast between the actual theory and practice of measurement, on the one hand, and the Copenhagen account of it, is summarized in Figure 2.10.

All the adherents of the Copenhagen interpretation agree that the apparatus is ever-present – a belief refuted by the fact that most of the problems solved in quantum theory concern exclusively quantons (Sect. 4.2). However, the Copenhagen camp is divided on the question of whether a quantum theory of measurement is necessary or even possible. Bohr denied both the necessity and the possibility, arguing that the general theory refers exclusively to experimental situations anyhow, and that the behavior of the apparatus has to be described classically. His favorite disciple conceded that a quantum theory of measurement is possible but insisted that it is unnecessary and warned that it is misleading: "a comprehensive quantal account of the measuring process (which is doubtless possible) is actually pointless (except inasmuch as it would provide a rather trivial check of the consistency of the quantum theory), since it would only obscure the essential function of the experimental arrangement in establishing the connexion between the quantal system and the classical concepts indispensable for its description" (Rosenfeld, 1964, p. 219).

A number of physicists, even while paying lip service to orthodoxy, have proposed a variety of general quantum theories of measurement. None of these theories has been put to the test, whence there is no objective ground for preferring any of them to the others. All these theories assume that the measurement results are both sharp and error-free, and they are so extremely general that they do not model any real measurement set-ups. We all know that every precision measurement involves some accidental error, and that there are no all-purpose measuring devices. Being so general, none of those theories allows one to design any measuring instruments or to calculate any precise and therefore testable predictions. In particular, they do not tell us how to measure positions and times, although the measurement of any property ultimately boils down to a sequence of position and time measurements. (Current quantum theory cannot solve satisfactorily the problem of position measurement because the position eigenstates are deltas, which are mathematically ill-behaved; the well-behaved position eigendifferentials correspond to position intervals, not points. As for time, it is a classical variable in quantum theory, so the latter has nothing new to say about it.) Such barrenness and remoteness from the laboratory of the general quantum theories of measurement is the more ironic given the apparent commitment of most physicists to operationism.

However, the utter failure of the *general* quantum theories of measurement does not condemn the program of building *specific theories* accounting for particular measurement processes in strictly physical (quantal *and* classical) terms. We need them to understand, and therefore operate correctly, and improve on, our laboratory equipment. (As a matter of fact there are already a number of such specific theories – e.g. of the blackening of photographic plates and of the working of the Geiger-Müller counter. Recall Sect. 1.1.) But we need a great many specific theories rather than a single fits-all theory. And the theories we need must not be purely quantal because all existing measuring instruments are classons, and quantum theory lacks the predicates required to represent some of their properties – e.g. rigidity, temperature, and viscosity. The theories we need are semiquantal or, equivalently, semiclassical. In particular they must account for the amplification process – e.g. in a photomultiplier – that starts in quantum events and results in macrophysical ones. And, of course, they must be realistic in the sense that they must treat the measurement process as a particular physical process occurring in a purely physical system – the measurement system described in Sect. 1.1. The theories of measurement that assume (rather than prove) that one can write and solve a Schrödinger equation for such a macrophysical system, and thus solve the measurement problem in all generality, are no better than theological doctrines.

Although we demand a multiplicity of specific theories of measurement – one for every type of measuring device – we admit that they share some features. In particular, there is nearly total consensus that they must account for the *projection* of the quanton state function (or "collapse" or "reduction" of the "wave packet") upon measurement. The intuitive idea is simple. (See e.g. Dirac 1958 §10.) If a measuring operation yields a "sharp" result, such as a single number, this must be an eigenvalue of the operator representing the property being measured; and, according to the eigenvalues postulate (Sect. 4.2), that eigenvalue must correspond to an eigenstate of the same operator (assuming non-degeneracy for the sake of simplicity). But then the original state of the quanton, which, according to the superposition principle, was a linear superposition of such eigenstates, must have been projected onto one of the axes of the functional space of the operator. In other words, the measuring apparatus filters out all the eigenstates except the one corresponding to the eigenvalue that is found upon measurement. In misleading pictorial terms: the initial wave packet, composed of many (perhaps all) eigenstates, reduces to one of them. See Figure 2.11.

Example 1 Polarization measurements, by means of a nicol prism in the case of light, or of a Stern-Gerlach apparatus in the case of massive quantons, actually *prepare* quantons in "sharp" states eigenstates), and so

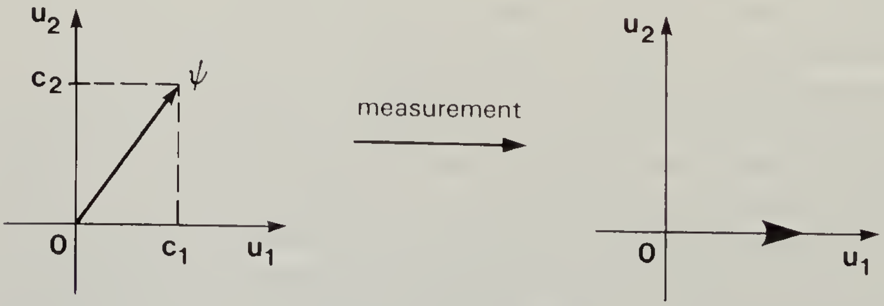


Fig. 2.11. The projection hypothesis: the initial state ψ of the quanton gets projected onto an eigenstate upon measurement (or, more generally, interaction): $\psi = \sum c_k u_k \rightarrow c_i u_i$. The operator effecting this projection is called a projector.

they effect projections. (Caution: only the state functions, not the properties themselves, get projected. For example, if the z-component of the angular momentum, or the spin, of a quanton acquires a definite value upon preparation or measurement, the other two components are blurred and therefore cannot be measured exactly. This follows from the non-commutativity of the angular momentum or the spin components.) *Example 2* Measurements of scattering cross sections involve projections at two stages: when preparing the (nearly) monochromatic incident beam, and when detecting the outgoing quantons. *Example 3* Some wavelength measurements end up in a reduction of the wave at the detector (e.g. the retina).

So far there is nothing problematic about all this except that it involves an idealization, for in many cases actual measurement results are numerical intervals rather than single numbers. (So, a more realistic version of the projection hypothesis is that measurement involves a projection of the original state function onto a subspace, rather than an axis, of the corresponding Hilbert space. But this is a mere technicality.) The real scientific and philosophical problems are posed by the *orthodox* version of the projection hypothesis (von Neumann 1932). Let us recall it briefly:

(i) As long as the quanton is not subjected to observation, its state function evolves smoothly according to the Schrödinger equation. This is called a *type II* process.

(ii) If the quanton is in state ψ before its "observable" (property) A is measured, the observation act throws ψ onto one of the eigenstates of the operator $\hat{A}$ representing A. This instantaneous, discontinuous and uncontrollable change is called a *type I process*.

(iii) It is in principle impossible to predict the precise eigenstate produced by an observation, but we can predict the probability of every possible outcome regardless of the type of measuring device. In fact the probability that ψ collapses onto the particular eigenstate u_k of $\hat{A}$ (Equation (5), Sect. 4.2) equals $|c_k|^2$, where c_k is the coefficient of u_k in the expansion of ψ into the eigenfunctions of $\hat{A}$.

(iv) The projection (reduction, collapse) is not a physical but an epistemological event: since ψ represents only our state of knowledge (not an objective state of affairs), its projection onto u_k represents the observer's change in his state of knowledge upon taking cognizance of the outcome of his observation.

In short, according to von Neumann and his followers there is projection of the state function upon measurement, but it is a purely subjective process consisting in the sharpening of the observer's information. In other words, *type I processes occur only in the observer's mind*: they do not affect the objects of measurement. (So much so that von Neumann restricted his theory to unobtrusive measurement. Only the observer-apparatus eigenstates were supposed to change as a result of the observation act.) This subjectivist interpretation of the state vector has been found wanting (Sect. 4.2). However, it allowed von Neumann to avoid the embarrassing questions of the mechanism and infinite speed of the "contraction": it did so by displacing them from physics to psychology. Nevertheless the orthodox version of the projection hypothesis posed more problems than it solved.

Firstly, the orthodox version *conflates apparatus and observer* – but we took care of this confusion a short while ago. Secondly, by assuming that the Schrödinger equation holds as long as nobody is observing, the orthodox view renders it *untestable*. (On top of this it contradicts the Copenhagen dogma *Thou shalt not feign any unobservables*.) Thirdly, it renders quantum probabilities *subjective*, and correspondingly turns mean standard deviations (those occurring e.g. in Heisenberg's inequalities) into *uncertainties* caused by the arbitrary or uncontrollable action of the observer-apparatus. (Ironically, this turns quantum mechanics into (*a*) a branch of epistemology, and (*b*) a causal theory – though one with unknowable causes.) Fifthly, by assuming that observation escapes the laws of physics, in particular the Schrödinger equation, the orthodox view treats measurement as an *unphysical* process – which contradicts the practice of measurement as well as the specific (and useful) theories of measurement. Sixthly, contrary to all laboratory experience, it assumes that measurement is *instantaneous*.

If none of the above objections has succeeded in shooting down the orthodox version of the projection hypothesis, think of the following contradiction generated by it (Everett 1973). Consider a quanton Q and an apparatus A by means of which some measurement is performed on Q. According to orthodoxy the projection hypothesis takes the place of the Schrödinger equation during the measurement process. Now consider a second apparatus B capable of observing the system $Q \dotplus A$ composed by Q and A. So long as B does not observe this composite system, the latter evolves smoothly according to the Schrödinger equation: it undergoes a type II process even while A is performing a measurement on Q. However, this contradicts the rule directing us to use the projection postulate to describe what goes on in the composite system $Q \dotplus A$. If on the other hand the second observer-apparatus B does perform a measurement on the composite system $Q \dotplus A$, how is he to describe this process? If $Q \dotplus A$ is treated as a physical system, a concession is made to realism, so the positivist philosophy inherent in orthodoxy is violated. And if the latter is respected, quantum theory leaves us in the lurch, for the projection postulate makes no provision for a second observer-apparatus – much less for a chain of observers, such that observer *N* observes what observer *M* is doing, who in turn observes observer *L*, and so on, down to observer *A*, who observes *Q*. So, in either case orthodoxy is untenable.

What is the theorist to do in the face of this disaster? If he is orthodox enough he will swallow the contradiction the way religious believers accept the mysteries of their religion. But if he is capable of independent critical judgment he has the following options: (*a*) to deny that projection occurs, *giving up* the projection hypothesis (Margenau 1936, Everett 1973, Fox 1983); or (*b*) to admit that projection occurs but try and *reform* the projection hypothesis, turning it into an approximate result derivable from the usual physical principles applied to a strictly physical process in which quantons interact with classons. The first solution is unpopular because it is generally recognized that projection does occur in measurement. The second solution is just as unpopular because it calls for the implementation of a difficult research project; therefore some people continue to employ the projection postulate in a pragmatic way, as an *ad hoc* heuristic rule (Lévy-Leblond 1977). We favor stand (*b*) for being principled and for suggesting an interesting research project; more on it in a moment.

(Everett's is the most original and oddest solution to the problem; it ought to appeal to those philosophers who like to play with possible worlds. According to him the state vector does not collapse upon measurement but splits into different branches or "worlds" – and this occurs continuously if the observation is continuous. These "worlds" would coexist and be mutually independent, i.e. non-interfering. But, because they are mutually independent, if we happen to inhabit one of them we will be unaware of all the remaining branches, so that everything *appears* to conform to the projection postulate. For this reason Everett's solution has been called "the many worlds interpretation of quantum theory": deWitt and Graham 1973. This solution to the contradictions generated by the orthodox version of the projection hypothesis is unscientific because the splitting is unobservable, so the conjecture is untestable. We are offered a piece of science fiction in place of a piece of defective science.)

We have been writing about the projection *hypothesis* where von Neumann and his followers wrote about the projection *postulate*. The reason is that we would like to see a rigorous *proof* that the projection, or something close to it, occurs partly as a consequence of the Schrödinger equation, not as a result of an arbitrary decision of an omnipotent Observer placed above the laws of nature. More precisely, we should like to derive a projection (or semiprojection) *theorem* from physical (quantal and classical) first principles. And we should like to have a proof that the projection (or semiprojection) is a *swift but not instantaneous process caused by certain interactions*, in particular those between quanton and apparatus. Such a proof would restore the unity of quantum theory and eliminate some of the unphysical features inherent in orthodoxy. The projection of a general state function onto one of its components should be just as natural and simple as the landing of a fair coin on one of its "eigenstates": while spinning in space the coin is 50% in the "heads" state and 50% in the "tails" state (Gottfried 1966).

A few physicists have attempted to accomplish just that: to prove projection from physical principles only. Thus Hepp (1972) proved that, in a highly idealized case, the state function of an individual quanton under observation gets reduced in the limit as time approaches infinity. (But, as Bell (1974) pointed out, "$t = \infty$ never comes, so that the wave packet reduction never comes".) More recently, Machida and Namiki (1980) have shown that projection is a result of micro-macro interactions, regardless of either amplification (required for perception) or irreversibility (inherent in recording). However, their treatment concerns whole ensembles of quantons rather than individual ones. Cini (1983) has tackled the problem in a far more realistic way, by studying specific measurements on a single quanton, namely with polarized counters and Stern-Gerlach devices. He shows that the "collapse" occurs to a high degree of accuracy, though not instantaneously, and as a consequence of the basic physical laws. (His treatment is confined to unobtrusive measurements.) Moreover the "collapse" may occur without irreversibility and, what is more important, it is a feature of micro-macro interactions in general, so it is bound to occur not only in the laboratory but throughout nature. This "eliminates that mysterious influence of the observer's mind on the microsystem behavior, which has been at the origin of so much frivolous nonsense by so many famous physicists" (Cini 1983 pp. 50–51). In short, progress is being made in demystifying the quantum theory of measurement and, in particular, in understanding the nature and mechanism of the state function projection.

The importance of research on the deduction and realist reinterpretation of the projection hypothesis can best be gauged by the paradoxes that its orthodox version generates. Only a moment ago we discussed the paradox constructed by Everett (1973). Another, even more striking, is the so-called *Zeno's quantum paradox*, according to which an unstable system subjected to continual observation does not change, much less decay (Misra and Sudarshan 1977, Chiu *et al.* 1977, Peres 1980, Horwitz and Katznelson 1983). Accordingly, Zeno would be right because Berkeley supposedly is. But at the same time the Copenhagen dogma, that whatever happens is the deed of some observer, would be refuted. Besides, we could all become immortal just by making sure that someone, other than God, is continually watching us.

The paradox stems from orthodoxy (Bunge and Kálnay 1983a, 1983b). It will be instructive to rehearse the paradox, trace its sources, and dissolve it. Consider an unstable quanton, such a free neutron or a radiactive atom, that can be in either of two states or in a superposition of both. Call those states the initial (or excited or undecayed) state, and the final (or ground or decayed) one. The probability that the system be in its initial (excited) state ϕ_1 at time t, either because it stayed in it or returned to it, equals the square of the projection of its state vector ψ onto the initial state ϕ_1, i.e.

$$p(t|0) = |(\phi_1, \psi(t))|^2. \tag{16}$$

For small values of t, a straightforward computation yields

$$p(t|0) \cong 1 - (\sigma/\hbar)^2 t^2, \tag{17}$$

where σ is the mean standard deviation of the quanton hamiltonian. Normally, to obtain the value $p(t|0)$ at any desired instant not too far off in the future one sets t equal to the desired number on the RHS of Equation (17), and no paradox arises.

To produce the paradox we must read Equation (17) in operational terms. Consider the unit interval [0, 1] and divide it into n equal subintervals, observing the quanton at the end of each of them by an unobtrusive means. The probability that the quanton be found in its initial state ϕ_1 at the end of the first time interval $1/n$ is, by Equation (17), $p_1 \cong 1 - (\sigma/n\hbar)^2$. Since by Equation (16) the "survival" probability does not depend upon the previous history of the quanton (hypothesis of statistical independence), after the second observation has been performed we shall have $p_2 \cong [1 - (\sigma/\hbar)^2]^2$. Finally, after the unit interval n/n, the probability of finding the quanton in its undecayed state is

$$p_n \cong [1 - (\sigma/n\hbar)^2]^n, \tag{18}$$

which approaches 1 as n approaches infinity. That is, the quanton stays in its initial state regardless of the length of the time interval, provided it is being observed all the time. The root of the paradox is obvious: every time an observation is made, even if it is an unobtrusive one, the state vector ψ of the quanton is assumed to project onto its initial state ϕ_1 in obedience to von Neumann's projection postulate. If the observations are made continuously, the quanton is prevented from changing because its state vector is forced to lie down on the initial state all the time: it cannot raise its head because Big Brother is watching. We proceed to show that Zeno's quantum paradox does not occur in our realist version of quantum theory.

By hypothesis the observation on our unstable quanton is unobtrusive (of the first kind). Hence it cannot affect the state of the quanton, which evolves spontaneously according to the Schrödinger equation. But, since the quanton is unstable, its initial state ϕ_1 cannot be an eigenstate of the energy operator: if it were, the quanton would not decay. The clue to the problem is the superposition principle, which tells us that, from the instant it is born – i.e. for all $t > 0$ – the unstable quanton is in a *superposition* of its undecayed (initial) state ϕ_1 and its decayed (final) state ϕ_2. In other words,

$$\psi(t) = c_1(\mathrm{t})\phi_1 + c_2(t)\phi_2, \tag{19}$$

where c_1 and c_2 are the projections of ψ onto the initial and the final states respectively, i.e.

$$c_i(\mathrm{t}) = (\phi_i, \psi(t)), \quad \text{for} \quad i = 1, 2,$$

and

$$|c_1(t)|^2 + |c_2(t)|^2 = 1 \quad \text{for all} \quad t. \tag{20}$$

As we saw in Sect. 4.2, $c_i(t)$ is the weight of the contribution of the sharp state ϕ_i, at time t, to the quanton state ψ.

Equation (20) suggests a simple geometrical representation that brings out the profound differences between quantons and classons. Pretending that the coefficients of the superposition (19), are real numbers, and using Equation (17), it is easily seen that the contributions of the sharp states evolve in time according to

$$c_1^2(t) = \cos^2 \omega t \cong 1 - (\sigma/\hbar)^2 t^2, \quad c_2^2(t) = \sin^2 \omega t \cong (\sigma/\hbar)^2 t^2. \tag{21}$$

The physical interpretation of these equations is obvious: see Figure 2.12. As time goes by, the state vector ψ rotates with angular velocity $\omega = \sigma/\hbar$ about the origin of the state space formed by the axes ϕ_1 and ϕ_2. In other

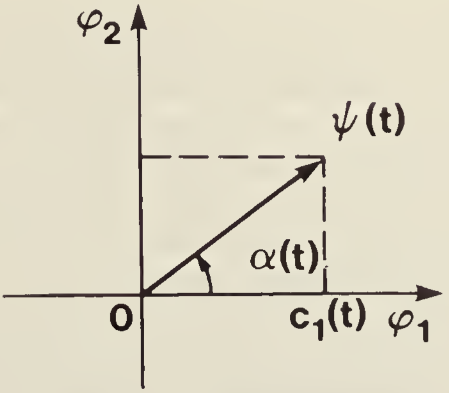


Fig. 2.12. Quantum Zeno's paradox. The unstable quanton is not always either in its initial (undecayed) state ϕ_1 or in the final (decayed) state ϕ_2. Most of the time it is in a linear combination of the two with time-dependent coefficients. The system goes back and forth between the two sharp states.

words, the quanton goes back and forth, in a continuous manner, between the two sharp states every $\hbar/\sigma$ seconds. In the classical case there is no rotation because the energy remains sharp at all times, i.e. $\sigma = 0$. As long as the quanton energy is spread ($\sigma \neq 0$), the quanton does not stay put in any energy eigenstate.

In conclusion, quantum Zeno's paradox does not arise if the projection hypothesis is not interpreted in a subjectivist manner, and the superposition principle is taken seriously. More on the second point in the next subsection (taken from Bunge 1984a).

6.2. *Hidden Variables, Separability, and Realism*

In the section we will discuss some developments that originated with the famous paper by Einstein, Podolsky and Rosen (1935) – henceforth EPR – and culminated with the experimental tests of hidden variables theories (in particular Aspect *et al.* 1981, 1982). In that paper it was held that the current quantum theory is incomplete for being about ensembles of similar things rather than about individual entities. (We criticized this view in Sects. 4.2 and 5.2.) On this view individual microphysical objects would have precise though different positions, velocities, energies, etc. Randomness would not be a basic mode of being but an appearance originating in individual differences among the components of an ensemble of mutually independent things. (Think of the distribution of bullet impacts around the bull's eye of a target.) A complete theory should contain hidden (classical) variables only, and it should deduce all the probability distributions instead of postulating them.

A hidden variable is a function possessing definite or sharp values all the time instead of being spread out: it is *dispersion-free* like the classical position and the classical field intensities. This type of variable was called 'hidden' in contrast with the so called "observables" and for allegedly referring to a sub-quantum level underlying that covered by quantum theory. (See Bohm 1957 for a fascinating account.) It was hoped that quantum theory could be enriched or completed with hidden variables, or even superseded by a theory containing only variables of this type. Bohm's theory (1952), which revived and expanded de Broglie's ideas on the pilot wave, was of the first kind. Stochastic quantum mechanics (de la Peña-Auerbach 1969) and stochastic electrodynamics (e.g. Claverie and Diner 1977) are of the second type. Moreover the latter theories, far from being thoroughly causal, contain random hidden variables like those occurring in the classical theory of Brownian motion.

Hidden variables theories were supposed to accomplish five or six tasks at the same time. It will pay to list and examine them separately because they are often conflated, with the result that there is still no consensus on what exactly experiment has refuted. Those tasks are:

(i) To *restore realism in the philosophical sense of the word*, i.e. to supply an objective account of the real physical world, rather than describe the information, expectations, and uncertainties of the knowing subject.

(ii) To *restore realism in EPR's idiosyncratic sense of the word*. That is, to comply with the oft-quoted EPR criterion of physical reality: "*If, without in any way disturbing a system, we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity*" (Einstein *et al.*, 1935, p. 777).

(iii) To *restore classical determinism* by deducing chance from causality – i.e. to explain probability distributions in terms of individual differences and the mutual independence of the components of statistical ensembles.

(iv) To *complete* the job begun by quantum theory, regarded as a statistical theory, by accounting for the (determinate) behavior of individual physical entities.

(v) To replace quantum theory , regarded as a phenomenological (or black box) theory, with a *mechanismic* theory that would explain quantum behavior instead of just providing successful recipes for calculating it. I.e., to exhibit the mechanisms underlying quantum behavior – e.g. electron diffraction and the tunnel effect –, deducing quantum theory as a particular case, in some limit, or for special circumstances.

(vi) To *restore the separability* or independence of things that, having been components of a closely knit system in the past, have now become widely separated in space – i.e. to eliminate the distant or EPR correlations. More on this below. Suffice it now to recall that, according to quantum theory, once a (complex) system, always a system (Vol. 3, Ch. 5, Sects. 4.1 and 4.2).

The first objective – the restoration of philosophical realism – was legitimate. However, it can be attained without modifying the mathematical formalism of quantum theory and, in particular, without introducing hidden variables. Indeed it can be shown that the Copenhagen interpretation is adventitious and that the realist interpretation, which focuses on things in themselves rather than on the knowing subject or observer, is not only possible but the one actually employed by physicists when not in a philosophical mood: recall Sects. 4 and 5. (For details see Bunge 1967c and 1973a.)

The second goal, though often mistaken for the first, has nothing to do with philosophical realism: it is simply *classicism*, and it involves the denial of the superposition principle, according to which the dynamical properties of a quanton normally have distributions of values rather than sharp values. (In the non-relativistic theory only the mass and the electric charge have sharp values all the time – but these are not dynamical variables.) In the light of the indisputable success of quantum theory and the failure of hidden variables theories, that classicist principle must be regarded as unjustified: we shall call it the *EPR dogma*. Paradoxically enough, this dogma can be upheld by surrendering classical logic. Indeed Friedman and Putnam (1978) have suggested that the variables of quantum mechanics can be regarded as "quantities [that] *always* have well-defined values" provided the distributive law for propositions is given up. This procedure has the advantage that it also restores realism in the philosophical sense (i). The catch is that the resulting theory bears only a superficial similarity to the quantum theory of physicists, which has ordinary logic built into it; in particular, this version of quantum theory yields the wrong probabilities: recall Sect. 5.2.

The third objective – determinism – was pursued by the early hidden variables theorists, but was no longer an aim of those who worked from about 1960 on. In fact, as we saw a while ago, there are theories containing random hidden variables, namely stochastic quantum mechanics and stochastic quantum electrodynamics. The introduction of these theories was important to distinguish the concepts of (i) philosophical realism, (ii) classicism or EPR realism, and (iii) determinism – as we realize with hindsight.

The fourth goal (completeness) makes sense only if quantum theory is regarded as a statistical theory dealing only with ensembles. It evaporates as soon as it is realized that the elementary theory refers to individual quantons (Sects. 4 and 5).

The fifth objective (mechanismic theory) was soon abandoned in the most advanced work on hidden variables: that of Bell (1964, 1966). In fact Bell's hidden variables have no precise physical interpretation: they are only adjustable parameters in a system that is far more phenomenological than quantum theory. Moreover such hidden variables are not subject to any precise law statements, hence they are no part of a physical theory proper. But this is a virtue rather than a shortcoming, for it shows that Bell's famous inequalities hold for the entire *family* of local hidden variables theories, regardless of the precise law statements that may be assumed.

So far, only the first objective – namely the restoration of philosophical realism – seems worth being pursued. However, it can be attained without the help of hidden variables, namely by a mere change of interpretation of the standard formalism of quantum theory. On the other hand the sixth objective, namely separability, remained plausible for about half a century. Indeed, the EPR distant correlations are quite counter-intuitive. How is it possible for the members of a divorced couple to behave as if they continued to be married, i.e. in such a manner that whatever happens to one of them affects the other? The quantum-theoretical answer is simple if sybilline: *there is no divorce*. However, let us not rush things: it will pay to have a look at the experiment suggested by Einstein *et al.* (1935). Here we shall deal with an updated version of it in the context of Bell's general hidden variables theory, which contains the now famous Bell's inequalities.

According to quantum theory the three components of the spin of a quanton do not commute, hence (in our interpretation) they have no definite or sharp values at the same time – as a consequence of which they cannot be measured exactly and simultaneously. Simpler: the spin components are not the components of a vector, hence there is no vector to be measured. On the other hand, if the spin were an ordinary vector (hence a hidden variable), its components would commute and consequently it should be possible to measure all three components at the same time with the accuracy allowed by the state of the art.

Apply the preceding considerations to a pair of things that are initially close together and with spins pointing in opposite directions, and subsequently move apart without being interfered with. Compute the probability $P(x, y)$ that thing 1 has spin in the positive x-direction, and thing 2 in the positive y-direction. According to quantum mechanics this probability exceeds the sum of the probabilities $P(x, z)$ and $P(y, z)$:

$$P(x, y) \geq P(x, z) + P(y, z). \qquad QM$$

On the other hand any hidden variables theory predicts the exact reversal of the inequality sign:

$$P(x, y) \leq P(x, z) + P(y, z). \qquad HV$$

This is one of Bell's inequalities. Bell himself and others derived several other inequalities, holding in all local hidden variables theories, and involving exclusively measurable quantities such as coincidence counter rates. (See Clauser and Shimony, 1978 for a masterly review.)

Thanks to this theoretical work a number of ingenious and highly accurate measurements were designed. Some of them, involving photons rather than massive quantons, have been performed. The latest version is shown schematically in Figure 2.13. An atom at S emits two photons at the same time, and their linear polarizations are measured by the polarizers I, oriented in the direction a, and II, oriented in the direction b. (These directions are externally adjustable.) The photons coming out of the polarizers activate the photomultipliers PM1 and PM2. The coincidence rate is monitored by the

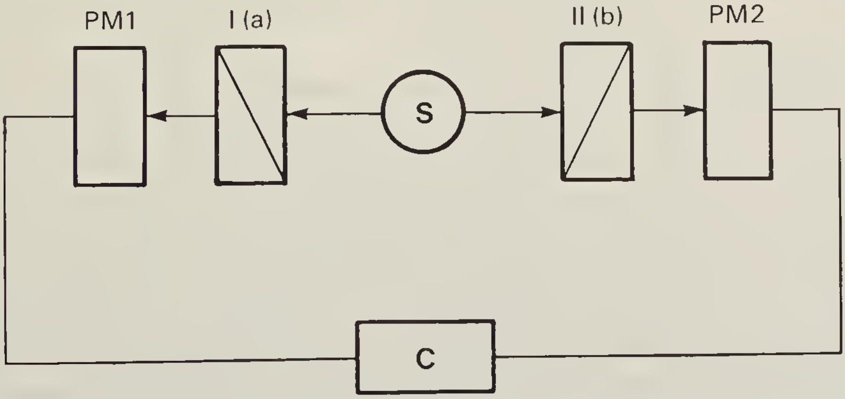


Fig. 2.13. Updated optical version of the EPR experiment. S: source. I(a) and II(b): polarizers (analyzers). PM1 and PM2: photomultipliers (amplifiers). C: coincidence counter activated only by the simultaneous arrival of signals from the two analyzers. Adapted from Aspect *et al.* (1982). Think not of two distinct things arriving at I and II but of a *single* system expanding up to I and II. The system breaks down when striking the photomultipliers PM1 and PM2.

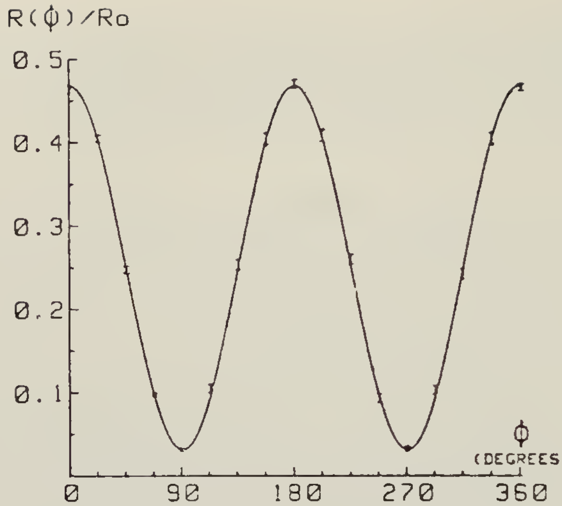


Fig. 2.14 The result of the EPR optical experiment diagrammed in Fig. 2.13. Coincidence rate as a function of the relative orientations of the polarizers. The data fit the quantum theoretical prediction (solid curve) almost to perfection. Taken from Aspect *et. al.* (1981).

counter C. The result is that the polarization states of the two photons are highly correlated: see Figure 2.14. That is, the state of polarization measured by polarizer II depends on that measured by polarizer I and conversely (Aspect *et al.* 1981). Essentially the same result obtains if the polarizers are rotated during the flight of the photons and before they strike the polarizers, to prevent any communication between them through signals propagating at a finite speed (Aspect *et al.*, 1982).

Although there is consensus that Bell's inequalities, and with them all of the hidden variables theories, have been experimentally refuted, it is not clear *what is at issue*: philosophical realism, determinism, the EPR dogma that all properties have sharp values at all times, or the EPR hypothesis that distant things behave independently of one another. (See Table 2.1.) Unfortunately most authors (e.g. Clauser and Shimony 1978, d'Espagnat 1979, Aspect *et al.* 1981) claim that *realism* is the casualty, without however clarifying what they mean by 'realism', and ostensibly conflating philosophical realism with what we have called 'classicism' or the 'EPR dogma'. Yet from our discussion in previous sections it should be clear that the very design and performance of measurements presuppose the reality (independent existence) of the entire measurement system, hence of every component of it (object, apparatus, and experimenter). So, philosophical realism is definitely not at issue: the downfall of the Bell inequalities has not refuted the principle that the physical world manages to exist without the help of those who try to know it. D'Espagnat (1980), when cornered by Michel Paty, has admitted this. So has Shimony (personal communication 1982).

TABLE 2.1. Terminological guide for the perplexed.

Expression	In this work	In the physics literature
system	system (thing with at least two components)	**1**: system **2**: physical object (simple or complex)
observer	observer	**1**: observer **2**: apparatus **3**: observer-cum-apparatus
realism	the thesis that nature exists by itself	the thesis that all properties have definite or sharp values all the time
classicism or EPR dogma	the thesis that all properties have definite or sharp values all the time	realism
hidden variables	dispersion-free functions	**1**: dispersion-free functions **2**: non-stochastic variables
Determinism	**1**: lawfulness plus *ex nihilo nil* **2**: causality	**1**: causality **2**: non-probabilistic lawfulness **3**: individual predictability
predictability	predictability	determinism
locality	nearby action	**1**: nearby action **2**: separability

Nor is determinism at issue, since some hidden variable theories entailing Bell-type inequalities contain random variables. (See e.g. Stapp 1980.) So, *only the EPR dogma and the separability (locality) hypothesis* are at issue. In fact this is the conclusion Clauser and Shimony (1978) reach after their careful analysis: Hidden variables (sharpness) & Locality (separability) ⇒ Bell's inequalities. Hence the experimental refutation of these inequalities refutes either separability or the EPR dogma – or both. We shall argue in a while that the *or* is inclusive, for the EPR dogma entails separability ("locality"). But before doing this we must have another glance at the Aspect *et al.* (1981, 1982) experiments.

The experiment poses two problems. One is to ascertain whether the quantons emitted by the source possess all their properties, in particular polarization, or acquire them when interacting with the polarizers. The second problem is to account for the correlations at a distance, or EPR correlations, between the results obtained with analyzer I and those obtained with analyzer II – or, equivalently, to explain the coincidences registered by the coincidence counter.

Let us start with the first problem. According to both classical and hidden variables theories, the fragments resulting from the process at the source have sharp properties (eigenvalues) from start to finish: the analyzers only exhibit what is already there. On the other hand, according to quantum theory the quantons leaving the source are in a superposition of polarization states. This superposition collapses (reduces) to eigenstates through interaction of the system with the analyzers. In other words, the analyzers do not just measure an existing polarization but also produce it: they are measuring instruments of the second kind. Clearly, this explanation will satisfy only those who believe that accurate measurement is accompanied by a reduction of the state function – even though one need not believe the dogmas that this reduction is subjective, instantaneous, and unrelated to the smooth processes described by the Schrödinger equation. (Recall Sect. 6.1.)

The second problem is this: If only one of the analyzers is read, we may infer the result obtained with the other analyzer without looking at it. (This cannot be done with the apparatus shown in Fig. 2.13, which registers only coincidences. The equivalent operation here is to alter the angle between the analyzers and note the change in coincidence rate. As shown in Figure 2.14, the latter attains its maximum when the two analyzers are parallel.) This result is nicely explained by any hidden variables theory jointly with the theorem of conservation of the total spin. Indeed, the first quanton would arrive at the analyzer possessing a definite spin value, say 1, so that the second quanton must have spin -1 since the total spin of the system at the source was 0). The trouble with this explanation is that it takes hidden variables for granted. Since experiment has refuted hidden variables theories we must try to explain the same results with the help of quantum theory.

If the two quantons leaving the source were to become independent after a while – as they should according to classical physics – the result obtained with analyzer I should be independent of the result obtained by analyzer II. Thus is should be possible to observe, say, the left quanton with its spin pointing in the x-direction, and the right quanton with its spin pointing in the z-direction. But this is not what quantum theory predicts: according to this theory there is total spin conservation, so that if the left analyzer projects the spin onto the positive x-axis, the right analyzer will project it onto the negative x-axis. In other words, there is a strong correlation at a distance between the polarization of the two former components of the original system. This correlation is correctly predicted by quantum theory, which treats the "two" quantons as *one*. (Only the state function for the entire system satisfies the Schrödinger equation.)

The quantum-theoretical explanation of distant correlation is often found unsatisfactory or even mysterious because it involves the counter-intuitive ("paradoxical") hypothesis that a complex system may continue to be such even after its components have moved far apart. In fact all known classical forces fall off quickly with increasing distance. For that reason a host of more or less unorthodox explanations of the EPR distant correlation have been proposed. Among them we note those in terms of hidden variables and a peculiar quantum potential that does not occur in the hamiltonian but derives from the state function of the system (Bohm and Hiley 1975), the presence of actions at a distance and even psychokinesis. We reject all three explanations for the following reasons. The first occurs in a hidden variables theory, hence one containing some of Bell's false inequalities. The second violates special relativity and electrodynamics (both classical and quantal), which involve the principle of nearby action (or locality in the field-theoretic sense). And the third violates the principle of conservation of energy, since every message must ride on a physical process – unless, of course, one is prepared to believe in ghosts.

We need not look for any special *mechanisms* explaining the distant or EPR correlations, any more than we need to explain the Lorentz length "contraction" and time "dilations" in terms of mechanisms (Paty 1981). Quantum mechanics accounts for distant correlations as follows. If two quantons are initially independent from one another, then the state function of the aggregate they constitute is correctly represented by the product of their individual state functions. So, separability or "locality" obtains in this case just as in classical physics. But if two quantons have initially been part of a system, i.e. if they have interacted strongly at the beginning, then the state function of the whole cannot be so factorized even after the components have moved far apart. In other words, the state of each component is determined not only by the local conditions (i.e. the state of the immediate surroundings of the spatially separated quantons), but also by their still belonging to a system. So, although physical separation entails spatial separation, the converse is false. Quantum theory is then *non-local* in this peculiar sense or, as we prefer to say, it is *systemic* in that it does not incorporate the classical principle that widely separated things cannot belong to the same system. (Note the ambiguity of the word 'local'. In field physics, whether classical or quantal, 'locality' means that nearby action, or non-action-at-a-distance, obtains: all actions are assumed to proceed *de proche en proche*. Quantum theory *is* local in this traditional sense of the word.)

We submit that the non-separability (or non-"locality") inherent in quantum theory is a consequence of the superposition principle together with the Schrödinger equation (Sect. 4.2). In fact, once a system is in a state consisting in an inextricable merger (not just either a sum or a product but a sum of products) of the states of its components, the Schrödinger equation guarantees that the system will evolve in time in such a way that its state function will keep that structural property even though the relative contribution of every individual eigenstate is likely to change in the course of time. (More precisely, consider a system with components I and II, and a property M that, like the z-component of the spin, can take only one of two values, m_1 or m_2, on each quanton at a time. Call u_1 and u_2 the states of component I in which M takes on the values m_1 and m_2 respectively, i.e. such that $Mu_1 = m_1u_1$ and $Mu_2 = m_2u_2$. Similarly, call v_1 and v_2 the corresponding states of component II. Then the total system $\mathrm{I} \dotplus \mathrm{II}$ is in some such state as $2^{-1/2}(u_1v_1 + u_2v_2)$ or $2^{-1/2}(u_1v_1 - u_2v_2)$. If the u's and v's are spin eigenstates, the preceding functions are distance-independent: the components continue to be entangled no matter how far apart they move.)

The EPR *paradox* is such only if one presupposes the *classical* principle that the behavior of things that are far apart cannot be correlated. (See Einstein, 1948 and Bell, 1966 for statements and discussions of this classical principle of "locality".) The physicist who takes this principle for granted

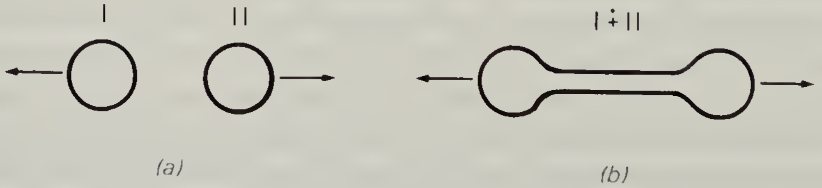


Fig. 2.15. The EPR experiment. (*a*) Classical account: two mutually independent things emerge from the source. (*b*) Quantum-theoretical account: the source emits a single system, the state function of which occupies a region that is in one piece.

is bound to puzzle over how it could be possible for two things to continue interacting after they have stopped being components of a system. Quantum theory solves the "mystery" much in the same way as a detective solves an alleged murder case by proving that there was no murder to begin with: that the system was not dismantled, that the distant components are still parts of the original system: see Figure 2.15. What the quantum theorist may puzzle over is a different question, namely: How, that being the case, could one effectively *dismantle* the system, and thus factorize its state function. We submit that the answer to this query is: The original system becomes dismantled only when at least one of its original components gets integrated into another system – e.g. when it is captured or absorbed by an atom.

In conclusion, (*a*) when two quantons interact, their state functions become entangled (not factorizable); (*b*) when the two quantons separate widely in space, they continue to form part of the original system although they do not act upon one another, much less at a distance and instantaneously (d'Espagnat 1981); (*c*) spatial separation is no cause for divorce: there is divorce only if there is new marriage; (*d*) non-separability is a consequence of the superposition principle and the Schrödinger equation; (*e*) non-separability is possibly "*the* characteristic trait of quantum mechanics" (Schrödinger, 1935b p. 555); (*f*) the failure of classical separability or "locality" (Einstein separability) confirms the systemic world view (Vol. 4) – not however the holistic one, because we do succeed in *conceptually* analyzing the composition and structure of systems; (*g*) in quantum theory there is EPR distant correlation (or EPR effect) but there is no paradox: the paradox arises only if quantum theory is combined with the classical principle of separability or "locality".

Philosophers have been quick to exploit the downfall of Bell's inequalities. Thus Fine (1982) has interpreted it as a refutation of determinism, and van Fraassen (1982) as a refutation of epistemic realism. We have seen that it is neither. In fact some hidden variables theories, such as stochastic quantum mechanics, are non-deterministic in that they contain random variables. (See additional criticisms by Nordin 1979 and Eberhard 1982.) As for epistemological realism, it would indeed be dead if, as van Fraassen contends, it involved the hypothesis that there is a causal mechanism underlying every correlation, since no such mechanism is known to exist in the case of the EPR distant correlations. But such a version of epistemological realism is a straw man. Epistemological realism happens to be an epistemological not an ontological view, and it boils down to the thesis that nature exists even if it is neither perceived nor conceived. (Consult any philosophical dictionary or recall Vol. 6, Ch. 15, Sect. 2.) Moreover, epistemological realism is compatible with alternative ontologies, in particular neodeterminism, which acknowledges non-causal (in particular probabilistic) types of determination (Bunge 1959a).

If anything, the experimental refutation of Bell's inequalities, like any other experiment, has confirmed epistemological realism once more, since every well designed experiment involves a clear distinction between knowing subject, object of knowledge, and apparatus. (Recall Sects. 1.1. and 6.1. See also Wheeler 1978 and Rohrlich 1983.) Only "realism" in the idiosyncratic sense of the EPR paper has been refuted along with Bell's inequalities – but that was only a classicist dogma. Actually this dogma went down the moment quantum theory succeeded in solving the problems that classical physics had solved incorrectly or was incapable of formulating. Since then we have learned that reality is smudged rather than neat: that every quanton possesses some properties that, far from having sharp values all the time, have distributions of values; and that, unlike rigid bodies but rather like fluids and fields, quantons are extremely sensitive to their environments, whether natural or artificial. This discovery did not alter nature and, in particular, it did not enslave nature to the observer. It only taught us that nature is composed not only of classons but also of quantons and semiquantons. It altered our representation of the world, not the world itself: it was a scientific revolution, not a cosmic cataclysm.

Hidden variables theorizing remained outside the mainstream of physics until Bell and others succeeded in deriving formulas capable of being subjected to rather direct crucial experimental tests. Then it came briefly to the fore, and it may soon become little more than a historical curiosity. However, the failure of hidden variables theories has taught us a valuable lesson: that, like other revolutionary theories before, quantum theory has got to be understood in its own terms. (See Bunge 1967c and 1973a.) Besides, hidden variables theorists have rendered a valuable service to the scientific and philosophical communities. Firstly, they have criticized some of the inconsistencies and obscurities of the Copenhagen orthodoxy. (Unfortunately they have often confused the issues – e.g. the problem of objectivity with that of determinism. See Bunge 1979c.) Secondly, they have helped others remove those inconsistencies and obscurities by reinterpreting the standard mathematical formalism of quantum theory in a strictly physical, hence realist, fashion. Thirdly, they suggested the very experiments that were to refute classicism. Now that the hidden variables episode seems to be closed for good we can look forward to theories even less intuitive than quantum theory. One wonders whether Einstein was right in believing that God (= Nature) is not malicious.

In a last ditch effort, the anti-realist might look elsewhere for support – and he is likely to find it in a number of carelessly worded texts. In particular he may claim that quantum theory postulates (*a*) a number of *unobservables* – and how could these bolster epistemological realism?; (*b*) that quantons of the same kind (e.g. electrons) are *indistinguishable* – and how can we make sure that there are two things out there if we are unable to distinguish them?; and (*c*) that quantons are nothing but the embodiments of *symmetries*, which would be Platonic ideas. Let us meet these objections.

It is true that quanton theory – like every other fundamental theory – includes a number of concepts lacking observational counterparts. Stationary states (in particular ground states) and virtual particles are cases in point. However, they are not of the same kind. The occurrence of the former only shows that quantum theory is unfaithful to operationism. (Realism holds that reality is necessary though insufficient for observability. Only positivism equates existence with observability.) As for the so-called virtual entities and processes, they are horses of a different color – or, rather, no horses at all. For example, two electrons are said to interact via a virtual photon, and two quarks via a virtual gluon. Such entities and processes cannot be real because they violate certain physical laws, notably energy conservation. Consequently they are admittedly unreal (virtual) and therefore unobservable – and yet they are postulated by scientists who pay lip service to operationism. Unlike stationary states, potentials, and other concepts with real counterparts, virtual things and processes are dispensable fictions (Bunge 1970). They have been invented only because some physicists, particularly Feynman, insist on attributing a physical (and if possible pictorial) interpretation to every term in a formula and every line in a diagram (Bunge 1955b). Only the naive realist demands a one-to-one correspondence between the concepts of a theory and "elements of reality" (an EPR expression). Scientific realism demands only that there be a whole theory-whole thing correspondence (Vol. 1, Ch. 3, and Vol. 6, Ch. 15). In short, unobservables, in particular virtual things and processes, can offer no comfort to the anti-realist.

As for indistinguishability, it is usually said that, if two particles of the same kind (e.g. electrons) are components of a system, then they are identical and therefore indistinguishable. (See Figure 2.16.) The unwary philosopher might conclude the refutation of Leibniz's principle of the identity of indiscernibles. But a careful analysis of the formulas in question

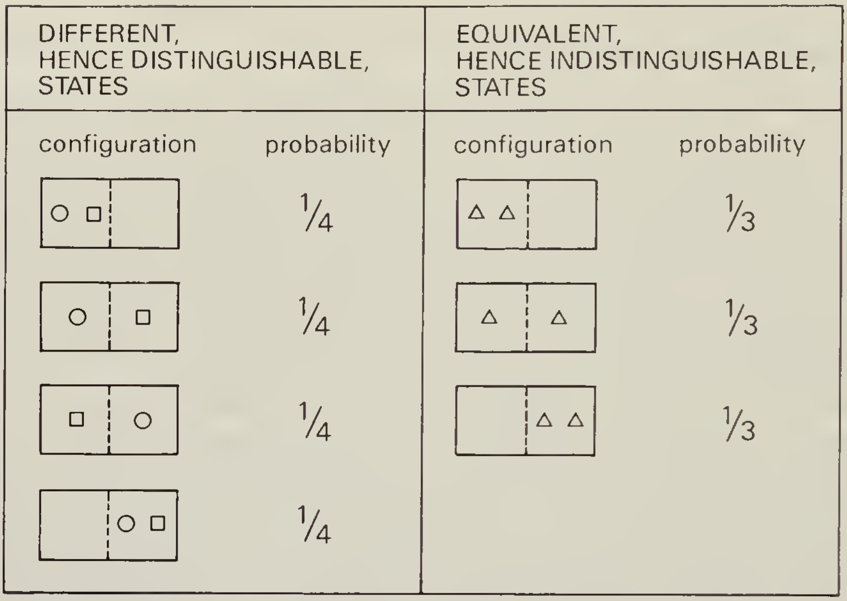

DIFFERENT, HENCE DISTINGUISHABLE, STATES
EQUIVALENT, HENCE INDISTINGUISHABLE, STATES
configuration
probability
configuration
probability
¼
¼
¼
¼
⅓
⅓
⅓

Fig. 2.16. "Indistinguishable" (exchangeable) components of a system. The state spaces of the system will differ or not according as the states "object 1 in cell 1 and object 2 in cell 2", and "object 1 in cell 2, and object 2 in cell 1", are assumed to be the same or different.

reveals that the physicist means *equal* or *equivalent* when he writes 'identical', and *exchangeable* when he says 'indistinguishable'. Indeed he assumes that the particles in question are different since he speaks of *two* or more entities, not of one. (When constructing the state function, e.g. the Slater determinant, of the system, he starts by writing 'N!' in the denominator, where N is the number of particles in the system – and he trusts that some experimenter will be able to count them, i.e. to determine the actual value of N.) What happens is that an exchange of such quantons makes no difference to the state of the system. In other words, the correct formulation of the principle is: "The state of a system of quantons of the same species is invariant under permutations of its components". Once again, no comfort for the anti-realist.

Finally, some physicists – notably Heisenberg (1969) – have stated that symmetry precedes existence: that quantons are nothing but embodiments of symmetries. This Platonic delusion stems from the manner the theorist confronts the bewildering array of "fundamental particles". Instead of proceeding inductively, or else by trial and error, he imagines that there is a single basic quanton that can be in different mass, charge, spin, isospin, hypercharge, etc., states. He then forgets for a while that hypothetical entity and investigates the algebraic properties of its state space, guided only by extremely general physical principles. In particular, the theorist investigates the group-theoretic structure of that space: he conjectures, say, that the structure is an US(2), US(3), or some other symmetry group. Once he is done he glances at the "particles" chart and, if lucky, discovers that nature does contain "embodiments" of such conjectured symmetries; and if very lucky, he will predict one or two still unknown "embodiments". But this is no evidence for the power of pure mathematics to mirror the world: all such symmetry groups are constructed on the basis of law statements such as commutation formulas for dynamical variables. Nor do symmetries hover over things, let alone produce them; in physics every symmetry is a property of either a physical entity, such as a crystal, or of a feature of a physical entity, such as the hamiltonian or the state function of a molecule. Melt the crystal, or dissociate the molecule, and the corresponding symmetry disappears.

In conclusion, quantum theory accounts for quantons as things quite different from classons, and therefore it offers a rather counterintuitive ("paradoxical") picture of reality. However, this picture is consistent, extremely rich and fertile, and perfectly compatible with epistemological realism. Inconsistencies, obscurities, paradoxes and uncertainties appear only when trying to force the theory into either the classical or the positivist mold.

7. CHEMISTRY

7.1. *Philosophy and Chemistry*

We define chemistry as the scientific investigation of chemosystems and their molecular products. (In turn, a chemosystem was defined in Vol. 4, Ch. 2 as a system wherein chemical reactions, such as $H + H \rightarrow H_2$, occurs.) Because chemosystems are composed of physical things, and are in turn often components of biosystems, chemistry lies between physics and biology. More precisely, physics helps explain chemistry, which in turn helps explain biology. This well known fact, along with the remarkable success of chemists in accounting for the marvelous variety of substances – millions of kinds of them – in terms of atoms of only about 100 kinds, and the stunning success of chemistry as the scientific basis of the powerful chemical industry, accounts for the popularity of chemistry since the mid-19th century.

Given the popularity and prestige of chemistry, it is strange that the corresponding philosophy hardly exists. The publications in this field are only a handful (e.g. Prélat 1947, Caldin 1961, Theobald 1976, Lévy 1979, Bunge 1982b and Vol. 4, Ch. 2, Primas 1983). Not even the distinguished philosophers of science Meyerson, Broad and Bachelard, who started out as chemists, made any significant contributions to the philosophy of chemistry: they preferred to write about other sciences. One cause for the underdevelopment of the philosophy of chemistry is that the vast majority of chemists are experimentalists impatient with foundational or even theoretical issues – to the point that theoretical chemistry constitutes a rather small and slowly growing body of knowledge. A second cause may be the philosophical misconception that chemistry raises no philosophical problems of its own for allegedly being but an application or even a logical consequence of physics.

Be that as it may, the philosophy of chemistry is nearly virgin soil that deserves being cultivated vigorously because it contains a rich and important problematics. The short and haphazard list that follows should confirm this statement. What can ontology learn from chemistry with regard to qualitative change, in particular emergence, and the intertwining of cooperation with competition in the chemical formation and breakdown of systems? Are the various laws of chemical reactions (around 6 million of which are known) mutually independent, or just instances of a single general law – or neither? How realistic are the ball and spoke models of molecules? Given that modern chemistry is based on atomic physics, does it follow that it is but an application of the latter, or does chemistry have concepts and hypotheses of its own? In this section we shall deal only with a small sample of the problematics of the philosophy of chemistry.

A first surprise that the chemical literature has in store for the philosopher is the frequency with which teleological and even anthropomorphic expressions occur in it. Examples: "The dissociation energy of a molecule is the energy *needed* to dissociate it", "An endothermal reaction is one that *requires expenditure* of energy to keep it going", "The energy that *must be supplied* to a molecule *in order to* break a bond is called the 'activation energy'", and "High energy bonds are enzymatically broken down to yield the energy *needed* to drive biosynthetic reactions".

Such expressions originate probably in the pragmatic attitude natural to the bench chemist and, particularly, to the industrial chemist. They are quite harmless in basic chemistry, for no chemist dreams of using them to explain anything: they are just sloppy descriptive phrases, and they have a legitimate place in chemical technology (industrial chemistry and chemical engineering). Moreover, if pressed any chemist is willing and capable of translating any such phrase into a rigorous scientific formula, e.g. "The dissociation energy of a molecule is the minimal energy that dissociates it". However, those teleological and anthropomorphic expressions can do great harm when exported to biology, which is still in the grips of ancient finalism. Whence a first task for the philosopher of chemistry: that of policing the language of the science.

Even deeper metaphysical trouble awaits the philosopher intent on examining the most innocent-looking chemical formulas, such as the familiar stoichiometric formula "NaCl" and the common reaction formula "$H + H \rightarrow H_2$. These formulas are of the types

$$AB \tag{22}$$

and

$$A + B \rightarrow C \tag{23}$$

respectively. In either case one may read them either platonically or nominalistically. In the former case one presupposes the reality of universals, in this case of chemical species; in the latter case one assumes that only individuals are (physically) real. Indeed, (22) may be read either as "Compound AB is composed of species A and B", or "Every individual of species AB is composed of one unit of species A and another of species B". Likewise (23) may be read either as "Species A and B combine to produce species C", or "One unit of species A combines with one unit of species B to produce one thing of species C".

In most cases this double talk is harmless. But when it comes to the foundations of chemistry the difference in interpretation will suggest alternative research lines and consequently different theoretical results. The Platonic interpretation suggests considering every chemical as the resultant of vectors in a vector space the base of which is constituted by the chemical elements (species), and a chemical reaction as a transformation of vectors of this kind (Aris 1965). On the other hand the nominalistic interpretation suggests forming a very different vector space, namely the state space of the system; here the axes are the concentrations of the various chemicals involved, be they elements or compounds (Vol. 4, Ch. 2).

Are chemical equations merely descriptive or do they explain as well? They are supposed to discharge both functions. (See Vol. 6, Ch. 10, Sect. 1.1 for the differences between these two operations.) However, chemical equations are only components of descriptions and explanations.

Strictly speaking, the mere existence of a compound or of a reaction calls for a whole special theory (or model). This is sometimes overlooked by experimental chemists, who are often content to describe a compound, find out its atomic components, and describe one of the reactions that produce it. Therefore the philosopher's task is in this case to stress the need for theories explaining what has been described.

Take for example a reaction of the form (23) that happens not to occur under normal pressure and temperature conditions unless a catalyzer is present. In the old times catalyzers, such as enzymes, where believed to do their job by just being present. We now explain the effect of a catalyzer in producing a reaction of type (23) by assuming that it occurs in two steps, each of which involves the catalyzer E, which is recovered at the end of the overall reaction. Indeed, we split (23) into two reactions:

$$A + E \rightarrow AE \tag{24a}$$

$$AE + B \rightarrow C + E. \tag{24b}$$

AE is an unstable complex (or rather a species of unstable complexes each composed of one unit of kind A and one of kind E). The above pair is called the *reaction mechanism*, which is supposed to explain the overall reaction. However, this only pushes the problem one step further. We will not have a full explanation as long as we do not know *why* the above intermediate reactions occur – i.e. why the complex AE is formed in the first place and why it combines with B to produce C. No doubt, postulating the intermediate reactions is justified by isolating the catalyst E and regulating the reaction velocities by varying its amount, as well as by isolating and analyzing the complex AE (provided it is not too short-lived). But the intermediate reactions themselves remain to be explained. Physical chemistry supplies this explanation at the macrolevel, particularly with the help of the concepts of affinity and dissociation energy. And quantum chemistry will one day be called upon to explain them at the deeper level of atoms and molecules, of which more below.

Another problem of growing interest to philosophy as well as chemistry is this: How universal a science is chemistry? Or, put in ontological terms: How widespread are chemosystems in the universe? So far chemistry has been mainly Earth-bound, and the few dozen chemical compounds found in other planets, in meteorites, and in interplanetary space, occur on our planet as well. However, the known chemosystems may prove to be only a smallish subset of the set of all existing chemosystems. Indeed the physical and chemical conditions on the surface of the explored planets are quite exceptional, in that the gravitational, radiation and temperature conditions are midway between those prevailing in the interior of stars and those prevailing in outer space. It is conceivable that at very high pressures, radiation densities, and temperatures, totally different chemical reactions and molecules occur. And it is possible that very short-lived molecules be formed in the interior of stars, in defiance of high temperatures but aided by intense gravitation. Of course this is speculative cosmochemistry. However, we should take such speculation into account before we adopt a geocentric view about the parochialism of chemistry.

To complete the chart of chemical compounds and reactions we need not only more experiments and more space exploration: we also need more theoretical chemistry for, after all, theory alone can tell us which compounds are possible and which ones are not. This platitude warrants repetition because, although chemists do not normally proceed by trial and error like alchemists, but rather are guided by hunches based on well-tried law statements, they do sometimes mistrust theory. Three theories are particularly promising at this time. One is still classical chemical kinetics, which is simple in that it involves only rate equations. However, these equations are non-linear and, moreover, they are systems of equations in the case of simultaneous reactions, like those occurring in most chemical reactors and in all living cells. In addition to such (very severe) computational difficulties, there is the methodological problem of measuring exactly the reaction velocities, which sometimes are huge. If only for its practical complications, good old chemical kinetics will always keep chemists occupied.

A second promising field of theoretical chemistry, and actually a more exciting one, is the merger of classical chemical kinetics with the physical theory of diffusion. This fusion, initiated by the mathematician Alan Turing in 1932, is of particular philosophical importance because it is a clue to morphogenesis, which, though usually described, is seldom explained. Indeed, the coupling of the two processes, chemical reaction and diffusion, gives rise to new patterns, such as propagating or standing chemical waves. Conceivably, such study of morphogenesis at the chemical level will inspire biologists interested in explaining biological morphogenesis. (Not that the process is likely to be the same, but that an explanation of it may come from a merger of previously separate disciplines.)

A third and even more exciting and promising field of theoretical research is physical chemistry. The goal of this study is to explain as much chemistry as can be explained with the help of physical theories such as thermodynamics, statistical mechanics, and quantum mechanics. An early triumph of quantum chemistry was the explanation of pair bonding with the help of Pauli's exclusion principle. (According to this postulate, any given state of a system of quantons of the same kind, e.g. electrons, can be occupied by at most two quantons. This principle explained the "empirical" generalization that the typical stable molecule contains an even number of electrons – a generalization that led G.N. Lewis to invent the concept of a pair bond as early as 1916) Other triumphs of quantum chemistry have been the explanation (at least in principle) of molecular spectra and the qualitative prediction of the existence of certain compounds. True, the number of exact results is small, and even the approximate results concern only a tiny fraction of all the kinds of molecule known to exist. However, the introduction of quantum mechanics into chemistry has effected a conceptual revolution. Even if they cannot calculate everything they would like to, chemists now think in a radically new way – e.g. in terms of potential barriers, ground and excited states, molecular orbitals, and scattering cross sections.

Quantum chemistry is expected to succeed where classical chemistry, in particular the classical electronic theory of valence, had failed – such as the existence of covalent bonds (as in H_2), of molecular spectral bands, and of individual chemical reactions (as in the collision of two atoms resulting in the formation of a molecule). But so far, around 99% of quantum chemistry has concentrated on the properties of light molecules: it is quantum chemical statics. Only a few papers deal with quantum chemical dynamics – the theory of chemical reactions –, which ought to become the heart of theoretical chemistry. Fortunately this field has gained vigor since about 1970, particularly since the method of crossed molecular beams was employed to study individual chemical reactions. (This is the direct method of altering the atomic composition of molecules. Classical chemistry knew only the indirect method involving myriads of atoms. Note the parallel with molecular genetics and classical genetics.)

The existence of quantum chemistry poses in new terms the old problem of reduction: Is chemistry merely a branch of physics? To be sure chemistry borrows many concepts from physics: witness those of atom, temperature, and specific heat. But it also mints concepts of its own, such as those of chemical reaction and catalyzer, affinity and reaction velocity. However, no sooner is a chemical concept proclaimed to be peculiar to chemistry, than someone attempts to reduce it to physical concepts. An early case was that of chemical purity, which pre-atomic chemistry was incapable of defining. (It gave *criteria* of chemical purity, such as the constancy of melting and freezing points, but no *definition* proper. The contemporary definition is: "Substance x is chemically pure $=_{df} x$ is composed exclusively of either atoms or molecules of a single species or kind".) A more recent example is the concept of chemical structure, which has been held to be alien to physics (Woolley 1978), but proved to be reducible to (definable in terms of) physical concepts after all (García-Sucre and Bunge 1981). Therefore nearly every quantum chemist harbors the secret hope that every single chemical concept will eventually prove to be defined in terms of physical concepts.

Yet the hard fact is that the successes of quantum chemistry have been qualitative rather than quantitative. One reason is admittedly that it involves tremendous computational difficulties, e.g. handling determinants with tens of thousands of entries. Yet it may also be the case that only the simplest problems in chemistry can be tackled *ab initio*, i.e. by posing and solving the Schrödinger equation. It may well be the case that medium and large molecules have emergent properties – such as a definite shape and the self-duplication of the DNA molecule – that are classical rather than quantal. In short, it may well be that many molecules – possibly all the 10^{12} or so kinds of protein, and the 10^{10} or so species of DNA – are semiquantons rather than quantons. If this is the case, *ab initio* calculations are foolhardy, and a semi-classical theory will have to be invented. This would be required not just because of the computational difficulties of purely quantum-mechanical models of molecules and reactions, but because medium and large molecules have emergent properties representable by concepts alien to quantum theory.

The radical reductionist is likely to rejoin that the problem of reduction is artificial for, after all, things and their properties do not come labeled 'physical' or 'chemical': they are said to be physical, or chemical, only because they happen to be conceptualized by physicists or chemists respectively – a division of labor that is conventional and subject to historical changes. Granted. However, it must also be owned that chemosystems and their molecular products have properties of their own (Vol. 4, Ch. 2), and that physics, though necessary to explain chemistry, is not enough. For one thing, there is the matter of the objective emergence of new properties, i.e. of properties possessed by systems but not by all of their components (Vol. 4, Ch. 1). For example, a body of water is not fully characterized by saying that its molecules are composed of two hydrogen atoms and one oxygen atom. The fluidity, transparency, low viscosity, low conductivity, and large dielectric power of water are macroproperties rooted to its composition but not understandable in terms of it alone. Witness the large differences between steam, liquid water, snow crystals, and ice cubes – not to speak of droplets and streams. For another, quantum mechanics does not seem capable of explaining molecular evolution. Not that this process is mysterious, but its explanation calls for not only a knowledge of atoms and molecules but also some knowledge of their interactions with a complex and variable environment that cannot be adequately represented by a Schrödinger equation because it is a process occurring at the interface between quantons and classons.

However, the problem of the reducibility of chemistry to physics deserves a new section, not only because of its intrinsic interest but also because its investigation may be used as a model for investigating similar problems, such as that of the reducibility of biology to chemistry, or of psychology to biology. In what follows we shall make heavy use of a previous study (Bunge 1982b).

7.2. *Is Chemistry Reducible to Physics?*

Opinion is divided as to whether chemistry is an independent science or a branch of physics. Experimental and classical theoretical chemists are likely to defend the autonomy of their science, pointing out that it has its own peculiar concepts, such as those of chemical bond and chemical reaction; its own peculiar law statements, such as reaction and rate equations; and its own peculiar experimental techniques, such as titration and chromatography – not to speak of its own professional societies and journals. On the other hand quantum theoretical chemists are likely to defend the opposite thesis, reminding us that even the old atomic theory explained the classical laws of chemical combination as well as the meaning of the stoichiometric formulas (such as "CO_2"); that the theory of molecules is but an application of quantum mechanics to systems composed of nuclei and electrons; and that the theory of chemical reactions is in principle an application of the quantum theory of scattering to the inelastic collision of atoms and molecules. If asked why, this being so, quantum chemistry has so far solved so very few problems, they may reply that it is all just a matter of computation ability: that "in principle", or "conceptually", the problem has been solved.

Who is right, the autonomist or the reductionist? To answer this question we need to compare, not just the referents of the theories of chemistry to those of physics, but all the aspects of the two fields of inquiry, We shall do so with the help of our characterization of a field $\mathcal{R}$ of scientific research

(Ch. 1, Sect. 1.1). This allows us to introduce a definition that will be found useful not only in this section but also in subsequent chapters:

DEFINITION 2.1 If $\mathscr{R}_i = \langle C_i, S_i, D_i, G_i, F_i, B_i, P_i, K_i, A_i, M_i \rangle$, for $i = 1,2$, denote two fields of scientific research at a certain time, then $\mathscr{R}_2$ is *included* in $\mathscr{R}_1$ (or $\mathscr{R}_2$ is a branch or chapter of $\mathscr{R}_1$) if, and only if (i) C_2 and S_2 are subsystems of the community C_1 and the society S_1 respectively, and (ii) every coordinate (component) of $\mathscr{R}_2$ beyond the second is included in the corresponding coordinate (component) of $\mathscr{R}_1$. (The inclusion relation is $\subseteq$.)

Let us check whether the ten relations occurring in the above definition hold among the components of physics ($\mathscr{R}_1$) and chemistry ($\mathscr{R}_2$). To begin with, it is empirically false that the chemistry community C_2 is a subsystem of the physics community C_1, even though some individuals work both as physicists and as chemists. Indeed each of those systems has its own professional and scientific organizations, and its own information network. On the other had, usually the society hosting C_1 and C_2 is one and the same – i.e. $S_1 = S_2$, the trivial case of the subsystem relation.

As for the domains, at first sight that of chemistry (D_2) is included in that of physics (D_1) because chemical systems would seem to constitute a special kind of physical systems. But this impression is mistaken, for what are physical about a chemical system are mainly its components, as the system possesses certain emergent chemical properties in addition to some resultant physical properties. Indeed, recall that a chemical system is a system wherein chemical reactions occur, i.e. one the atomic and molecular composition of which is variable. Since system and components do not share all their properties, they do not belong to the same species of thing. In short, it is not true that $D_2 \subset D_1$; rather, the composition of a chemosystem is included in D_1

On the other hand, the philosophical background G_2 of chemistry is properly included in that of physics. In particular, for most purposes chemists adopt an atomistic or corpuscularistic ontology: field-theoretic ideas enter only through quantum theory. As for the formal backgrounds F_1 and F_2, they may be taken to be identical. It is true that chemistry makes little use of certain mathematical tools, such as differential geometry, that field physicists find indispensable. However, this restriction may be temporary: it is advisable to have the totality of mathematics on call.

Divergences reappear from the sixth coordinate on. To begin with it is not true that the specific background B_2 is included in B_1. Rather, chemistry draws liberally from physics, i.e. $B_2 \subset K_1$. But, since physics employs a modicum of chemical knowledge, we also have $B_1 \cap B_2 \neq \emptyset$. (For example, solid state physicists need to know the chemical composition of the substances they study.) In sum, it is not true that $B_2 \subset B_1$ and $K_2 \subset K_1$. As for the problematics of the two sciences, they overlap partially; but, since chemistry studies a domain of facts that is not included in the domain of physics, it is not the case that P_2 is included in P_1. The general aims of chemistry are the same as those of physics, namely description, explanation, and prediction; however, since chemistry studies facts that are not covered by physics (i.e. $D_2 \not\subset D_1$), the specific aims of chemistry are not included in those of physics: $A_2 \not\subset A_1$. Finally, for this very reason $M_1 \not\subset M_2$. In sum, we have proved that *chemistry is not included in physics*.

However, there is consensus that the study of chemosystems *depends* upon a knowledge of physical entities: one also says that the former is *based* on the latter. What is not always clear is the very concept of epistemological dependence of one field of inquiry upon another. Fortunately all the elements for an elucidation of this notion are at hand. A first obvious condition for the dependence relation between two fields of inquiry to hold is that they have something in common, for otherwise neither could help the other. A second condition is that the specific background B_2 of the dependent field be included in the fund of knowledge K_1 of the other. A third condition is that the very formulation of every problem in the dependent field require the solution to some problems in the other. We assume that these three conditions are necessary and sufficient for the dependence relation to hold. And we lay down the following formal definition for ease of reference:

DEFINITION 2.2 The research field $\mathcal{R}_1 = \langle C_1, S_1, D_1, G_1, F_1, B_1, P_1, K_1, A_1, M_1 \rangle$ *precedes epistemologically* (or is presupposed by, or serves as a *basis* for) the research field $\mathcal{R}_2 = \langle C_2, S_2, D_2, G_2, F_2, B_2, P_2, K_2, A_2, M_2 \rangle$ at a certain time if and only if, at that time,

(i) $\mathcal{R}_1$ and $\mathcal{R}_2$ *intersect*, i.e. at least one of the coordinates of $\mathcal{R}_1$ beyond the second has a non-empty overlap with the corresponding coordinate of $\mathcal{R}_2$;

(ii) the specific background B_2 of $\mathcal{R}_2$ is *included* in the fund of knowledge K_1 of $\mathcal{R}_1$;

(iii) *every* problem in P_2 is implied by *some* problem(s) in P_1 – i.e. formulating and investigating any problems in P_2 requires the previous solution to some problem(s) in P_1.

It is easy to see that physics precedes chemistry in this sense, just as mathematics precedes physics, and chemistry precedes biology. Indeed, condition (i) is satisfied because $G_2 \subset G_1$ and $F_2 \subseteq F_1$; condition (ii) is satisfied as well because chemists make use of, *inter alia*, thermodynamics and quantum mechanics; and condition (iii) too is satisfied because the very formulation of every chemical problem presupposes having solved certain physical problems, such as measuring weights or calculating entropies. So, *chemistry is based on physics* in the sense of Definition 2.2.

We shall support our conclusion by examining the relation between quantum chemistry and quantum physics, the more so since doing quantum chemistry looks very much like doing quantum physics – so much so that the logo of the Quantum Chemistry Group (Uppsala, Sweden and Gainesville, Florida) is Ψ. We shall do so by using the concepts of concept reduction and theory reduction. To start with the concepts of quantum chemistry: whereas some of them (e.g. those of atom and state function) are borrowed from the quantum theory of atoms, others are not. We say that a chemical concept B is *fully reducible* to physics if it is definable exclusively in terms of a (simple or complex) physical concept A, i.e. if $B =_{df} A$. The concepts of valence and molecular structure, though typically chemical, are definable in terms of the quantum theory of atoms. On the other hand the very concept of molecule is not so definable. Indeed a molecule is a whole described by a state function the properties of which are determined not only by its physical components but also by their interaction. So much so, that a molecular state function is not built by combining the state functions of the constituent atoms, but is found by solving the Schrödinger equation for the molecule. Likewise the concept of rate constant, central in the theory of chemical reactions, is not fully reducible to physics even though the value of the rate constant(s) of any chemical reaction can in principle be calculated with basic quantum mechanics plus certain subsidiary assumptions.

(The concept of rate constant is defined in *classical* chemical kinetics by reference to the phenomenological rate equations. The quantum-theoretical calculations cannot even start without explicit reference to the corresponding classical rate equation(s). For example, the rate of formation $\dot{n}_C$ of the reaction product C in a bimolecular reaction of the form "$A + B \rightarrow C$" is given classically by $\dot{n}_C = kn_A n_B$, where n_X is the concentration of chemical of kind X. An aim of quantum chemistry is to *calculate* k, defined classically, in terms of the probability of collisions of A's with B's, instead of leaving it as an unexplained empirical parameter. See Polanyi and Schreiber (1973) for the theory, and Bunge (1982b) for a detailed methodological analysis.)

The concepts of molecule and of rate constant are cases of partial concept reduction. This methodological concept can be defined as follows. Let A be a physical concept, B a chemical one, and C a chemical proposition not deducible from physical propositions alone, such that "If C, then $B = A$". We call this a *partial reduction* of B to physics.

As for *theory* reduction, we shall use the definitions introduced in Vol. 6, Ch. 10, Sect. 3.1. Let $\mathscr{T}_1$ and $\mathscr{T}_2$ be two theories the domains or reference classes of which have a nonempty intersection. Further, call Δ a nonempty set of reductive definitions (or full concept reductions), and A a nonempty set of hypotheses not included in either $\mathscr{T}_1$ or $\mathscr{T}_2$ but couched in the language of $\mathscr{T}_1$. Then (i) $\mathscr{T}_2$ is *fully* (or *strongly*) *reducible* to $\mathscr{T}_1$ iff the union of $\mathscr{T}_1$ and Δ entails $\mathscr{T}_2$; and (ii) $\mathscr{T}_2$ is *partially* (or *weakly*) *reducible* to $\mathscr{T}_1$ iff $\mathscr{T}_2$ follows logically from the union of $\mathscr{T}_1$, Δ, and A. Moreover, we say that $\mathscr{T}_1$ *helps explain* $\mathscr{T}_2$ iff $\mathscr{T}_1$ is mechanismic and $\mathscr{T}_2$ phenomenological, and $\mathscr{T}_2$ is fully or partially reducible to $\mathscr{T}_1$.

Our discussion of the concept of rate constant suggests that quantum chemistry explains classical chemistry, but that it is only *weakly reducible* to the quantum theory of atoms. We shall sketch a proof of the latter assertion. The question is to ascertain whether the set A of subsidiary assumptions occurring in the preceding definitions is empty, i.e. whether or not quantum chemistry follows from quantum mechanics without further ado. It is easy to pinpoint members of A; only, some of them look so obvious after half a century that theorists hardly bother to state them explicitly, and therefore fall victims to the illusion that quantum mechanics entails chemistry. Among such subsidiary assumptions we count the following:

A_1 Every molecule is composed of nuclei and electrons. [Note the difference with the classical atomic theory, according to which atoms are the units of chemical analysis.]

A_2 All the interactions among the components of a molecule are electromagnetic [even in the case of homopolar bonds].

A_3 Every molecule has an equilibrium state which is that of lowest energy (i.e. its ground state).

A_4 Every chemical reaction is either elementary or the outcome of a number of simultaneous elementary processes of the same kind.

A_5 Every chemical reaction consists in the combination, dissociation, or substitution of atoms or polyatomic systems, such as molecules and radicals. [This hypothesis is borrowed from classical chemistry.]

In addition to such extra substantive hypotheses one also introduces a number of methodological ones. The strongest among these are simplifications such as neglecting the motion of the nuclei, or assuming this or that electron to be far away from the nuclei, or even neglecting whole components of the molecule, such as the sugar-phosphate backbone of DNA. Moreover, most quantum chemical calculations involve guessing zero approximation state functions. Even more to the point: many calculations in quantum chemistry, particularly in chemical dynamics, involve classical statistical mechanics and even classical billiard ball models (Levine and Bernstein 1974, Miller *et al.* 1976). In such cases the quantum-chemical models are semiclassical. Only for comparatively simple molecules and chemical reactions can we speak of the partial or weak reduction of quantum chemistry to the quantum theory of atoms. (See also Lévy 1979, Primas 1983.)

To sum up, (i) whereas some chemical concepts are definable in terms of physical concepts, others are not; (ii) the quantum theory, by itself, fails to entail quantum chemistry: the latter follows only when enriching the former with extra assumptions, so that quantum chemistry is based on quantum physics but is not part of it; (iii) chemistry, regarded as a field of inquiry, has a number of traits which it does not share with physics – hence chemistry is not a branch of physics; (iv) nor is chemistry an autonomous science, for it is based on physics. However, in turn chemistry feeds physics a number of data, ideas, and problems. So, chemistry and physics are *interdependent.* The explicit recognition of this interdependence has always benefited both sciences. On the other hand isolationism, or purism, has caused stagnation or at least missed opportunities.

8. MEGAPHYSICS

8.1. *Earth Science*

In this section we shall glance at the physics and chemistry of planets, in particular our own. That is of course earth science, which comprises geology, geochemistry, geophysics, oceanography, and meteorology. So far, earth science has been the object of only a few philosophical studies, mostly by practitioners of the science (e.g. Albritton 1963, Kitts 1974, 1977, Riccardi 1977). However, it poses a rich and interesting philosophical problematics, as shown by the following sample. Is planetary science exclusively observational or also experimental? Is it only descriptive or also explanatory and predictive? Does the fact that it is a historical science alienate it from physics and chemistry? Does it have concepts, hypotheses and methods of its own, or are they all reducible to physical or chemical ones? And what has been the impact of the plate tectonics revolution, initiated in the 1960s, on the methodics of earth science? Let us examine a few such problems.

There are two rather popular misconceptions of the nature of earth science (Kitts 1974). One is that it is an exclusively *observational* discipline not an experimental one. Although this description does apply to most of the work in the field, it fails to cover all of it. In fact, geologists and seismologists make experiments on models, and meteorologists experiment with miniclouds in ice boxes, as well as with natural clouds (e.g. by seeding them with crystals). The difference between earth science and the other factual sciences, with respect to experiment, is one of degree and phase of development, not one of kind (Riccardi, 1977).

A related popular belief is that earth science is an exclusively *descriptive and inductive* discipline, one that proceeds from case to case and from sets of similar cases to timid empirical generalizations. This belief too is false: earth science is not any more inductive than other sciences. Far from restricting himself to descrining what he sees, the earth scientist frames or borrows hypotheses to explain his data – in particular, hypotheses about the unobserved evolution of the planet. And, since most of the hypotheses he borrows from physics belong to hypothetico-deductive systems, he often engages in deductions.

Another received view is that, because earth science is undoubtedly a *historical* discipline, it must be radically different from physics and chemistry (Simpson 1963). This opinion is only partly true. Firstly, earth science has its synchronic aspect as well. Secondly, only the land masses have long term memory: the atmosphere and the oceans (unlike the ocean floors) have only short term memory, so it is impossible to trace their history except indirectly, via the land masses. So, earth science is only *partly* historical – just like astronomy. Thirdly, it is not the case that physics and chemistry are totally non-historical. In fact, physics studies materials with structural or magnetic memory: recall metal fatigue and the common magnetic tape. And the study of molecular evolution, so important to biology, is a growing chapter of chemistry. Ergo, physics and chemistry too are historical, though only peripherally so for the time being. But even if the historical portions of physics and chemistry were excised, earth science would not be alienated from physics and chemistry. If it were it would be impossible to use physical and chemical laws to account for the evolution of the planet. In short, earth science is *both* historical and firmly based on physics and chemistry (Kitts 1974).

That earth science makes constant use of physical and chemical law statements, is well known. For this reason it satisfies one of the conditions for being ranked as a science (Vol. 6, Ch. 14, Sect. 2.1). Besides, earth science has a number of generalizations expressing regularities in the evolution of the planet – e.g. Steno's stratigraphic generalizations such as the "law of superposition". But, since these generalizations do not belong to full-fledged hypothetico-deductive systems, they do not qualify as law statements in our rather restrictive sense (Vol. 5, Ch. 9, Sect. 3.2). In sum, at least for the present earth science has no laws of its own: it borrows laws from physics and chemistry. Thus seismology is an application of elasticity theory; meteorology, one of hydrodynamics and thermodynamics; and geochemistry, of chemistry.

Two reasons may be adduced for the absence of laws peculiar to earth science. One is that it is a comparatively young science. However, this won't do because there are even younger sciences, such as psychology, that boast of a few laws. A deeper reason may be that there *are* no objective and universal historical patterns of the evolution of planets – or of anything else for that matter. After all, a history is not a law but the *result* of laws and circumstances (e.g. initial composition, boundary, and constraints). In other words, a law is a bundle of possible trajectories in some state space, rather than a single trajectory (Vol. 5, Ch. 9, Sect. 3.2). This being so, the very search for historical laws, in earth science or elsewhere, is "mistaken in principle" (Simpson 1963 p. 29). All we can hope to find is definite *trends* in a given history.

At first sight earth science has not even concepts of its own: it would seem that it contains only specializations of physical and chemical concepts, as well as peculiar names to designate them. Thus geological strata and tectonic plates form special kinds of bodies, seismic waves are a particular case of elastic waves, and 'Earth' is just a name of a particular planet, planets being only a kind of mechanical object. True, but such specializations are *unique* to earth science: they do not occur in basic physics or chemistry, and they make sense only in earth science. Thus the concepts of rock and sand, of mountain and river, of karst and volcano, of aurora borealis and earthquake, of wind and rain, are characterizable only with the help of a comparatively large number of physical or chemical concepts each. In sum, earth science *does* have concepts of its own.

The recent plate tectonics revolution has not altered the nature of geology. For one thing, it has been as respectable a scientific revolution as any in that, far from cutting all links with the past, it has grown out of the previous period, it retains many of its achievements, and it keeps physics and chemistry as constraints not to be violated. Thus the plate tectonics revolution does not quite exemplify Kuhn's original catastrophic view (Kitts, 1977). Rather, it illustrates that blend of evolution and revolution we all call 'evolution' (Vol. 6, Ch. 13, Sect. 3.2).

The most dramatic effect of the plate tectonics revolution has been to catalyze the clumping of the branches of earth science that evolved nearly in parallel. (See e.g. Smith Ed. 1981.) In turn, this synthesis has enhanced considerably the explanatory and predictive powers of earth science. In fact geology, which had focused attention on gradual processes, such as sedimentation and erosion, is now capable of explaining at least some of the catastrophes occurring in the terrestrial crust. The explanatory power of the theory can be gathered from what I take to be its five postulates: (i) The Earth's crust is composed of a few plates (at latest count 15 shells about 100 km thick each). (ii) The plates are torsionally rigid and they float on the mantle. (iii) The plates fit rather closely together. (iv) The plates are in slow motion, sometimes rubbing against each other and at other times creeping underneath their neighbors (subduction). (v) The frictions at the plate boundaries generate seismic waves that propagate throughout the Earth's crust.

Much remains to be known about the plates, in the first place the kind of forces that drive them. However, even in its present young stage, plate tectonics theory has contributed powerfully to explaining the present configuration of mountain ranges and ocean bases, as well as many earthquakes. It also explains continental drift, which had been laughed out of academia when first proposed by Wegener in 1919. (Wegener's failure to propose a mechanism for the drift, along with the dogmatism with which gradualism was adhered to, was the undoing of his hypothesis – until it was resurrected.) In turn, continental drift explains such orographic processes as the formation of the Himalayas as resulting from the collision of India against China. It also solves such biogeographical puzzles as certain commonalities in the flora and fauna of Africa and South America, which had once been united. And it may also explain certain remarkably quick speciations and extinctions in the history of life on our planet. In short, the hypothesis of continental drift explains a large mass of previously disconnected and unexplained data. (The latter constitute indirect or circumstantial evidence for the hypothesis, which so far has not been confirmed by direct measurement because of the slowness of continental drift, estimated to be on the order of a few centimeters per year.)

Plate tectonic theory focuses on catastrophes, but these are not the only events in the history of the planet. There are also gradual processes, such as sedimentation and erosion, that constitute the continuous background occasionally interrupted by quick violent processes. In this regard earth science is no different from any other historical science: both gradual and catastrophic changes have their legitimate places. Hegelians might say that this is a dialectical synthesis between the traditional catastrophism of the "successive creations" and gradualism. Nothing of the sort, for ancient catastrophism had been miraculous rather than scientific, and because the catastrophes envisaged by contemporary earth scientists (as well as by evolutionary biologists) are really processes – only, far quicker than the others.

Plate tectonics theory has renewed some disciplines, such as paleogeography, the task of which is to reconstruct the distribution of land masses in the past – and perhaps even to forecast their future distribution. In turn, paleogeography has shed intense light on evolutionary biology, and it is likely to alter some ideas in paleoanthropology. Last, but not least, that revolution has led to the emergence of a *new system* of knowledge out of a number of previously rather disconnected disciplines. The new system is, of course, earth science.

In conclusion, there is no question that earth science, though firmly based on physics and chemistry, is a distinct science. In fact it is not only cultivated by a distinct scientific community, but it has concepts of its own (e.g. that of geological stratum), problems of its own (e.g. describing and explaining changes in landscapes), methods of its own (e.g. surveying with the help of Landsat satellites), and aims of its own (e.g. reconstructing the history of our planet and locating mineral resources). All it needs, to catch up with the rest of the physico-chemical sciences, is more imagination of the kind that led to formulating the plate tectonics theory (and its precursor, Wegener's continental drift hypothesis) and the design of the Mohole project. A realistic philosophy of earth science should contribute to stimulating that imagination.

8.2. *Cosmology*

Cosmology was born thousands of years ago in close association with religion and philosophy. (See e.g. Munitz 1957.) It acquired something resembling scientific status in the 19th century, when it was first conceived as the mechanics and thermodynamics of the universe. Since then it has undergone two revolutions. The first was brought about in the 1920s by the application of Einstein's theory of gravitation to the universe as a whole. The second revolution started about 1970 and is still in progress: it resulted from an uneasy merger of Einsteinian cosmology with "particle" physics – whence the name "quantum cosmology".

Each of the three transformations fired the popular and literary imaginations, but only the first made a deep impression on philosophers. Yet contemporary cosmology poses a large number of interesting philosophical problems, such as: Is cosmology possible given that it deals with a single thing and that there is no science but of the general? Are there cosmic laws? Could the universe have been created, and if so was it designed as the home of man? If not, what evidence could there possibly be obtained about the stage preceding the "big bang"? And Is the universe bound to end up in a state of homogeneous disorder? (See also Kanitscheider 1979, Harrison 1981.)

Let us start with the first question. It has often been claimed that cosmology cannot be scientific because it concerns a single entity, whereas laws, which are essential to science, are universal. This view rests on too narrow a conception of natural laws. To be a law statement a formula need not be referentially universal: it suffices that it be spatially or temporally universal. For example, the schema "For all times thing *b* satisfies condition *L*" is admissible as a law schema. With this proviso it is clear that the formulas of cosmology, such as the Robertson-Walker line element, will become law statements the day they are empirically confirmed. When this happens they will not differ in general structure from the laws of motion of the Moon, or of Mars, which are unique to those bodies. On the other hand there is room to doubt that the Hubble "law" ($v = H_0 r$), which relates the velocity of a moving light source to its red (or blue) shift, is a law proper. It may be merely an empirical generalization expressing a trend in the observed part of the universe: it may prove to be false in regions yet to be explored or in future epochs. After all, at the time of writing the data are consistent not only with the hypothesis of monotonical expansion but also with that of an eventual contraction.

Another crucial problem still to be solved is that of the spatial finitude or, alternatively, infinity, of the cosmos. The metric tensor (or line element) shared by most contemporary cosmological models, *viz.* Robertson–Walker's, contains a constant k that must be determined empirically on the strength of measurements of the average matter density in the universe. If $k > 0$, the universe is closed (spatially finite), whereas if $k \leq 0$, the universe is open or flat (spatially infinite). Current data are so imprecise that they are compatible with the two alternatives. Until this crucial point is resolved, the said metric cannot be called a *law*: it is a speculative conjecture.

However, even if cosmology had no laws of its own, i.e. even if all of its hypotheses were speculative, it would meet our definition of a science (Vol. 6, Ch. 14, Sect. 2.1) insofar as it uses laws borrowed from other fields – in particular the laws of gravitation and of "particle" physics. In this regard cosmology does not differ from its humbler yet more prosperous neighbor, earth science. Indeed both are *predominantly historical sciences based on physics* (as well as on chemistry in the case of earth science).

Granted, cosmology is still largely speculative. But speculation can be sound or wild, scientific or unscientific, according to whether it harmonizes, or fails to harmonize, with well-founded general hypotheses, i.e. law statements (Bunge 1983e). Hence any failure of a particular cosmological model to agree with the physics of the day can be used to indict it as unscientific. For example, steady state cosmology – the only serious rival of the big bang model between 1950 and 1965 – became unscientific the day it incorporated the *ad hoc* hypothesis of the continuous creation of matter out of nothing (Bunge, 1962b).

Nineteenth century cosmology yielded plenty of food for philosophical thought. The Kant-Laplace hypothesis on the origin and evolution of nebulae (our galaxies) became a pillar of the naturalistic and evolutionary world view. On the other hand the Newtonian model of the universe suggested the idea of eternal recurrence (actually invented in Antiquity). If the universe is a huge (perhaps infinitely large) gas in a steady state, then any given event is likely to recur any number of times. In particular – perish the thought – this very page may have been written before, and may be written again, by a perfect replica of the present author. This boring model was soon replaced by a pessimistic one deriving from an application of the second law of thermodynamics to the universe as a whole. It was suggested that the universe is running downhill: that its total energy, though quantitatively conserved, is becoming degraded. It was forecast that the universe will eventually end up in a state of maximum entropy or "thermal death", where no macrophysical novelty could ever happen.

All three models were eventually modified or even discarded. The Kant-Laplace hypothesis is too simple to be true: it ignores quantons. (Yet, it must be admitted that we have no good alternative theory of the origin of galaxies, much less one compatible with the fashionable big bang model.) On the other hand the Newtonian model of the universe as a whole, as a sphere of uniform density, has been rendered compatible with the expansion hypothesis and it yields a remarkably good approximation. As for the thermodynamic version of eschatology, it has been given up because the second principle holds only for closed systems, and the universe is not enclosed in a container. (Remember Sect. 2.2.) Besides, astronomers have discovered plenty of anentropic (or build-up) processes, such as the birth of new stars and galaxies, along with break-down processes.

Since about 1920 nearly all cosmological models have been compatible with Einstein's theory of gravitation. The most popular of them are the *evolutionary* or Friedmann models, which involve the Robertson–Walker line element mentioned a while ago, and assume that the constant k occurring in it is non-negative (i.e. that space is closed). The best known of all these models is the "standard hot big bang model", according to which (*a*) initially the universe was concentrated in a comparatively small "fireball" at about 10^{12} degrees Kelvin; (*b*) the ball exploded for some unknown cause about 2.10^{10} years ago, and the universe has been expanding ever since; (*c*) at the present time matter and radiation are distributed homogeneously throughout the universe.

Hypothesis (*a*) implies that the universe is spatially finite and, together with hypothesis (*b*), that its present radius is about 2×10^{10} light years. This is sometimes interpreted in theological terms, as proving that the universe was *created* out of nothing. Thus Jastrow (1978 p. 12): "the astronomical evidence proves that the Universe was created twenty billion years ago in a fiery explosion". This statement contains three misleading words: 'evidence', 'proves', and 'created'. Firstly, as we all know, no empirical evidence can prove definitely a hypothesis – particularly if the latter is out of step with the naturalistic world view inherent in modern science. Secondly, the available evidence supports also alternative models; more on this point below. Thirdly, the word 'created' is out of place in cosmology: all the standard hot big bang model does assume is that the universe *started to expand* about 2×10^{10} years ago. Moreover the model assumes explicitly that the universe existed prior to that time, though in an entirely different state, namely as a "fireball". Hence it is a gross distortion to hold that "the astronomical evidence leads to a biblical view of the origin of the world" (Jastrow 1978 p. 14). Astronomy does not assume that the universe was "born", let alone "created", when the fireball exploded. If it did it would be unscientific, for science abides by the principles that nothing comes out of nothing or turns into nothingness (Vol. 3, Ch. 1, Sect. 1.3), and that everything happens according to law rather than miracle (Vol. 3, Ch. 4, Sect. 2.3). In short, it is not true that contemporary cosmology supports creationism, and that science is at one with theology, as Jaki (1974), Jastrow (1976), Davies (1983) and others have claimed. Natural theology cannot be revived at this late date.

The great methodological question concerning the standard hot big bang model is the quality and quantity of the data in support of it – particularly since only the favorable evidence is usually cited. This evidence consists of the following items: (*a*) the model includes the best theory of gravitation we have, namely Einstein's: (*b*) the model predicted correctly the existence of the 3 °K background radiation, coming and going in all directions, and supposed to be a relic of the primeval explosion (as well as a universal reference frame); (*c*) the model explains the redshift of the spectra of nearly all distant galaxies. Point (*a*) is important but not crucial, because many alternative models are compatible with the same theory of gravitation. Point (*b*), the most astonishing prediction of the model, is its war horse. I submit, on the other hand, that point (*c*) is its Achilles's heel, for alternative explanations of the redshift are conceivable – and because it is shared by all expanding models. Let us pause for a moment and glance at the redshift controversy.

The redshift is a well known fact: what is at issue is its explanation. Theorists analyze the redshift into three components; (*a*) Doppler effect (due to the mutual recession of the galaxies), (*b*) gravitational (due to the action of gravitational fields on light), and (*c*) cosmological, or due to the expansion of the universe (wavelengths stretch as space itself expands). By framing special hypotheses and using certain additional observational data, astronomers can often ascertain the relative contributions of these three redshift sources. This analysis is then anything but a matter of routine – so much so that component (*c*), usually estimated by subtracting the sum of components (*a*) and (*b*) from the total observed value, should be handled with great caution.

The upshot of the preceding discussion is this. Since nobody can *see* the galaxies receding (Doppler effect) or space expanding (cosmological redshift), astronomers try to explain ("interpret") the observed redshifts with the help of well known physical laws, speculative hypotheses, and additional data. Neither of the last two are fool-proof. (The sun is the exception: here only the gravitational redshift is important and is known with accuracy. On the other hand the quasars, which exhibit huge redshifts, have been a source of spirited argument ever since they were discovered by the Burbidges: see e.g. Field *et al.*, 1973.)

Given the difficulty of the subject there should be nothing surprising about this controversy: there is no growing science without controversy. What is anomalous is the fact that most astronomers do not seem to recognize the controversial nature of the matter, but adopt the standard model uncritically and, "knowing that their subject is this area already rests on rather shaky foundations as far as hard-proven evidence is concerned, cannot face up to the opening of Pandora's box in extragalactic astronomy" (Burbidge 1981). In this situation a detailed methodological examination of the redshift controversy would prove useful for an objective evaluation of the existing cosmological models as well as to facilitate the inception of new ideas. However, there will be room here only for a few remarks. (See Tolman 1949 and Bunge 1955c for criticisms of the dogmatism with which the early version of the big bang model was adopted.)

The defenders of the standard hot big bang model assert that it accounts for *all* the known facts. This is not quite true. Firstly, the model assumes that the universe is spatially finite (closed), although the known values of matter density at present favor an infinite (open) model (Liang and Sachs 1980). Secondly, the model treats matter as if it were distributed continuously and homogeneously throughout the universe, and so it collides head-on with the most conspicuous observational feature of the universe, namely its inhomogeneity or local clumping into stars, star clusters, galaxies, etc. (Neyman 1963). Thirdly, the hypotheses about the first few nanoseconds of the expansive phase are brilliant speculations, but it is still to be proved that they are consistent with Einstein's theory of gravitation, which is about classons, not quantons. (It is far from obvious that a matter tensor built with the help of quantum theory, as demanded by "quantum cosmology", results in a meaningful geometric tensor. The union of general relativity and quantum theory should involve the quantization of the gravitational field, and this is still a goal to be achieved.) Fourthly, according to the standard model, space should be rather strongly curved; but astronomical observation suggests that, although space curvature does occur in the vicinity of large bodies, space in the large is very nearly flat – this being why some Newtonian cosmological models are a good approximation. Fifthly, the largest estimated intergalactic distance is about twice the distance (in light years) the age the model attributes to the expansive phase of the universe. These and other difficulties have prompted a number of astronomers to look for alternatives, among which the so-called hierarchical models (involving many levels of clumping) stand out.

In sum, although the standard cosmological model does account for some of the known facts, it is inconsistent with others. And even if it did account for all the data we should be cautious in accepting it , because the data are neither as numerous nor as accurate as to exclude alternative models. After all, many a cosmological hypothesis has proved to be no better than myth clad in beautiful mathematics (Alfvén 1977). The myth of the creation of the universe is one of them. A related myth is the "anthropic hypothesis", according to which the universe has been designed as the home for man (Carter 1974). According to this wild speculation the various fundamental physical constants would have values and relations that, far from being contingent, would point to the emergence of man at a suitable place and time. This exercise in neo-Pythagoreanism gives only the order of magnitude of the said constants and it lacks predictive power (Carr and Rees 1979). Moreover the logic is wrong, for from the fact that man emerged at the place and time he did, and not somewhere else at a different time, all that follows is that it was physically possible, not necessary, for our species to appear there and then.

To conclude. Cosmology is currently thriving as never before but, at the same time, it risks entering a barren phase for the premature adoption of one of the various models compatible with both fundamental physical theory and astronomical data. It is in dire need of a strong dose of critical realism.

9. Concluding remarks

Physics and chemistry pass for being the "hardest" of all sciences. Yet some of their areas – particularly "particle" physics, quantum chemistry, and cosmology – are far less developed than certain areas of biology, such as genetics, or even social science, such as history. Even the most successful physical theories contain grave unsolved difficulties, such as the infinities and virtual entities of quantum electrodynamics. Nor is there any doubt that the foundations of physics and chemistry are remarkably immature, partly because so few scientists work on them. But there is an additional cause of this backwardness, namely the operationist philosophy still adhered to by most physicists and chemists although it was refuted long ago by philosophers. Therefore philosophers have a chance and a moral obligation to meddle with the foundations of physics and chemistry – provided of course they are reasonably well acquainted with them.

We have only sampled the problematics of the philosophy of physics and chemistry. The following short and haphazard list of open problems should persuade anyone that the field needs many more laborers. What kind of science is meteorology: basic or applied, homogeneous or mongrel? Does the quantum theory of solids exemplify strong (full) or weak (partial) reduction? What is the root of the "arrow of time"? Is it true that quantum theory has rendered ontology impossible by welding the object to the cognitive subject? Are genuine elementary building blocks of nature likely to exist? Is it reasonable to blame so many unexplained astronomical anomalies on such conjectured entities as superluminal "particles" and black holes? We could go on and on, but we must stop here.

In Part II we shall jump from physics and chemistry to the life sciences. However, we shall meet again, under a different guise, some of the problems encountered in the present chapter, in particular those of anthropomorphism and reduction. But of course we shall also encounter some new problems – which would not be the case if biology were a mere continuation of physics and chemistry.

BIBLIOGRAPHY

Agassi, J., and R.S. Cohen, Eds. (1982). *Scientific Philosophy Today: Essays in Honor of Mario Bunge.* Dordrecht-Boston: Reidel.

Albritton, C.C., Jr., Ed. (1963). *The Fabric of Geology.* Stanford: Freeman, Cooper and Co.

Alfvén, H. (1977). Cosmology: myth or science? In W. Yourgrau and A.D. Breck Eds., *Cosmology, History and Theology*, pp. 1–14. New York: Plenum.

Anderson, A.R. and N.D. Belnap Jr. (1975). *Entailment.* Princeton: Princeton University Press.

Angel, R. (1980). *Relativity: The Theory and its Philosophy.* Oxford: Pergamon.

Arbib, M. (1977). Review of three books on fuzzy set theory and its applications. *Bull. Amer. Math. Soc.* **83**: 946–951.

Aris, R. (1965). Prolegomena to the rational analysis of systems of chemical reactions. *Arch. Ratl. Mechanics and Analysis* **19**: 81–99.

Arruda, A. I. (1980). A survey of paraconsistent logic. In A. I. Arruda, R. Chuaqui, and N. C. A. da Costa, Eds., *Mathematical Logic in Latin America* pp. 1–41. Amsterdam: North-Holland.

Aspect, A., P. Grangier and G. Roger (1981). Experimental test of realistic local theories via Bell's theorem. *Phys. Rev. Letters* **47**: 460–463.

Aspect, A., J. Dalibard, and G. Roger (1982). Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Letters* **49**: 1804–1807.

Axelrad, D. R. (1983). *Foundations of Probabilistic Micromechanics.* Oxford: Pergamon Press.

Ballentine, L. (1970). The statistical interpretation of quantum mechanics. *Rev. Mod. Phys.* **42**: 358–381.

Bar-Hillel, M. (1979). On the subjective probability of compound events. *Organizational Behavior and Human Performance* **9**: 396–406.

Barwise, J., Ed. (1977). *Handbook of Mathematical Logic.* Amsterdam: North-Holland.

Bell, J. and M. Hallett (1982). Logic, quantum logic and empiricism. *Phil. Sci.* **49**: 355–379.

Bell, J. S. (1964). On the Einstein-Podolsky-Rosen paradox. *Physics* **1**: 195–220.

Bell, J. S. (1966). The problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.* **38**: 447–452.

Bell, J. S. (1974). On the wave packet reduction in the Coleman-Hepp model. CERN preprint TH 1923.

Bellman, R. E. and L. A. Zadeh (1977). Local and fuzzy logics, pp. 105–165. In J. M. Dunn and G. Epstein, Eds. *Modern Uses of Multiple-Valued Logic.* Dordrecht-Boston: Reidel.

Benacerraf, P. and H. Putnam, Eds. (1964). *Readings in the Philosophy of Mathematics.* Englewood Cliffs, N.J.: Prentice-Hall.

Bergson, H. (1907). *L'évolution créatrice.* Paris: PUF, 1948.

Bernays, P. (1976). *Abhandlungen zur Philosophie der Mathematik.* Darmstadt: Wissenschaftliche Buchgesellschaft.

Beth, E. W. (1959). *The Foundations of Mathematics.* Amsterdam: North-Holland.

Birkhoff, G. and J. von Neumann (1936). Logic of quantum mechanics. *Annalen der Mathematik* **37**: 823–843.

Bishop, E. (1967). *Foundations of Constructive Analysis.* New York: McGraw-Hill.

Bishop, E. (1975). The crisis in contemporary mathematics. *Historia Mathematica* **2**: 507–517.

Blokhintsev, D. I. (1964). *Quantum Mechanics.* Dordrecht: Reidel.

Bôcher, M. (1905). The fundamental conceptions and methods of mathematics. *Bull. Amer. Math. Soc.* **11**: 115–135.

Bohm, D. (1951). *Quantum Theory.* Englewood Cliffs, N.J.: Prentice-Hall.

Bohm, D. (1952). A suggested interpretation of the quantum theory in terms of "hidden variables". *Phys. Rev.* **85**: 166–179; 180–193.

Bohm, D. (1957). *Causality and Chance in Modern Physics.* London: Routledge & Kegan Paul.

Bohm, D., and B. Hiley (1975). On the intuitive understanding of non-locality as implied by quantum theory. *Founds. Phys.* **5**: 93–109.

Bohr, N. (1934). *Atomic Theory and the Description of Nature.* Cambridge: Cambridge U. Press.

Bohr, N. (1935). Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.* **48**: 696–702.

Bohr, N. (1936). Kausalität und Komplementarität. *Erkenntnis* **6**: 293–303.

Bohr, N. (1949). Discussion with Einstein on epistemological problems in atomic physics. In Schilpp, Ed. pp. 199–241.

Bohr, N. (1958). *Atomic Physics and Human Knowledge.* New York: Wiley.

Born, M. (1953). Physical reality. *Phil. Quart.* **3**: 139–149.

Born, M. (1971). *The Born-Einstein Letters.* New York: Walker & Co.

Bourbaki, N. (1970). *Théorie des ensembles.* Paris: Hermann.

Bridgman, P. W. (1949). Einstein's theories and the operational point of view. In P. A. Schilpp, Ed., pp. 333–354.

Brouwer, L. E. J. (1975). *Collected Works*, Vol. 1, A. Heyting, Ed. Amsterdam: North-Holland.

Browder, F.E. and S. MacLane (1978). The relevance of mathematics. In Steen, Ed., pp. 323–350.

Brunschvicg, L. (1929). *Les étapes de la philosophie mathématique*, 3rd ed. Paris: Presses Universitaires de France.

Brush, S. G. (1983). *Statistical Physics and the Atomic Theory of Matter.* Princeton, NJ: Princeton University Press.

Bunge, M. (1951). What is chance? *Science and Society* **15**: 209–231.

Bunge, M. (1955a). Strife about complementarity. *Brit. J. Phil. Sci.* **6**: 1–12, 141–154.

Bunge, M. (1955c). *La edad del universo.* La Paz: Laboratorio de Física Cósmica.

Bunge, M. (1959a). *Causality.* Cambridge, MA: Harvard University Press. 3rd rev. ed.: *Causality in Modern Science* (New York: Dover, 1979).

Bunge, M. (1959b). *Metascientific Queries.* Springfield, Ill.: Charles Thomas.

Bunge, M. (1961). Laws of physical laws. *Am. J. Phys..* **29**: 518–529.

Bunge, M. (1962a). *Intuition and Science.* Repr.: Westport, Conn.: Greenwood Press, 1975.

Bunge, M. (1962b). Cosmology and magic. *Monist* **44**: 116–141.

Bunge, M. (1964). Phenomenological theories. In M. Bunge, Ed. *The Critical Approach: Essays in Honor of Karl R. Popper*, pp. 234–254. New York: Free Press.

Bunge, M. (1966). Mach's critique of Newtonian mechanics. *Am. J. Physics* **34**: 585–596.

Bunge, M. (1967). *Foundations of Physics*. New York: Springer-Verlag.

Bunge, M. (1968). Physical time: the objective and relational theory. *Phil. Sci.* **35**: 355–388.

Bunge, M. (1970). Virtual processes and virtual particles: real or fictitious? *Intern. J. Theoret. Phys.* **3**: 507–508.

Bunge, M. (1971). A mathematical theory of the dimensions and units of physical quantities. In Bunge Ed. pp. 1–16.

Bunge, M. (1973a). *Philosophy of Physics*. Dordrecht: Reidel.

Bunge, M. (1973b). *Method, Model and Matter*. Dordrecht: Reidel.

Bunge, M. (1974a). *Treatise on Basic Philosophy*, Vol. 1: *Sense and Reference*. Dordrecht: Reidel.

Bunge, M. (1974b). *Treatise on Basic Philosophy*, Vol. 2: *Interpretation and Truth*. Dordrecht: Reidel.

Bunge, M. (1974c). The relation of logic and semantics to ontology. *J. Philosophical Logic* **3**: 195–210 Repr. in Bunge 1981a pp. 175–192.

Bunge, M. (1976a). Possibility and probability. In Harper & Hooker, Eds., Vol. III, pp. 17–33.

Bunge, M. (1976b). The relevance of philosophy to social science. In W. Shea, Ed., *Basic Issues in the Philosophy of Science* pp. 136–155. New York: Neale Watson Academic Publ.

Bunge, M. (1977a). *The Furniture of the World*, Vol. 3 of the *Treatise*. Dordrecht and Boston: Reidel.

Bunge, M. (1977b). Quantum mechanics and measurement. *Int. J. Quantum Chem.* XII, Suppl. **1**: 1–13.

Bunge, M. (1979a). *A World of Systems*, Vol. 4 of the *Treatise*. Dordrecht-Boston: Reidel.

Bunge, M. (1979b). Relativity and philosophy. In J. Bärmark, Ed., *Perspectives in Metascience* pp. 75–88. Göteborg, Regiae Societatis Scientiarum et Litterarum Gothoburgensis, Interdisciplinaria 2.

Bunge, M. (1979c). The Einstein-Bohr debate over quantum mechanics: Who was right about what? *Lecture Notes in Physics* **100**: 204–219.

Bunge, M. (1980a). *Epistemología*. Barcelona: Ariel. French transl.: *Epistémologie* (Paris: Maloine, 1983). German transl.: *Epistemologie* (Mannheim: Bibliographisches Institut, 1983).

Bunge, M. (1980c). *The Mind-Body Problem*. Oxford: Pergamon. German transl. with foreword by B. Kanitscheider: *Das Leib-Seele Problem*. Tübingen: Mohr, 1984.

Bunge, M. (1891a). *Scientific Materialism*. Dordrecht-Boston: Reidel.

Bunge, M. (1981b). Four concepts of probability. *Applied Mathematical Modelling*. **5**: 306–312.

Bunge, M. (1982a). The revival of causality. In G. Fløistad, Ed., *Contemporary Philosophy* **2**: 133–155. The Hague: Nijhoff.

Bunge, M. (1982b). Is chemistry a branch of physics? *Zeits. f. allgemeine Wissenschaftstheorie* **13**: 209–223.

Bunge, M. (1983a). *Exploring the World*, Vol. 5 of the *Treatise*. Dordrecht and Boston: Reidel.

Bunge, M. (1983b). *Understanding the World*, Vol. 6 of the *Treatise*. Dordrecht and Boston: Reidel.

Bunge, M. (1983c). *Controversias en física*. Madrid: Tecnos.

Bunge, M. (1983e). Speculation: wild and sound. *New Ideas in Psychol.* **1**: 3–6.

Bunge, M. (1984a). Hidden variables, separability, and realism. *Rev. Brasil. Física* 150–168.

Bunge, M. (1984c). La necesidad de mantener la dicotomía entre verdades de razón y verdades de hecho. *Rev. Latinoamer. Filos.* **10**: 63–69.

Bunge, M. (1985). Seven types of rationality. In J. Agassi and I. C. Jarvie, Eds. *Rationality: The Critical View.* The Hague: Nijhoff.

Bunge, M., Ed. (1967a). *Delaware Seminar in the Foundations of Physics* New York: Springer-Verlag.

Bunge, M., Ed. (1971). *Problems in the Foundations of Physics.* New York: Springer-Verlag.

Bunge, M. and A. Kálnay (1983a). Solution to two paradoxes in the quantum theory of unstable systems. *Nuovo Cimento* **77B**: 1–9.

Bunge, M. and A. Kálnay (1983b). Real successive measurements on unstable quantum systems take nonvanishing time intervals and do not prevent them from decaying. *Nuovo Cimento* **77B**: 10–18.

Burbidge, G. (1981). Origin of red shifts. *Science* **213**: 1198.

Caldin, E. F. (1961). *The Structure of Chemistry in Relation to the Philosophy of Science.* London: Sheed and Ward.

Carnap. R. (1950). *Logical Foundations of Probability.* Chicago: university of Chicago Press.

Carr, B. J., and M. J. Rees (1979). The anthropic principle and the structure of the world. *Nature* **278**: 605–612.

Carter, B. (1974). Large number coincidences and the anthropic principle in cosmology. In S. M. Longair, Ed., *Confrontation of Cosmological Theories with Observational Data*, pp. 291–298. Dordrecht: Reidel.

Casanova, G. (1947). *Mathématiques et matérialisme dialectique.* Paris: Ed. Sociales.

Chiu, C. B., Sudarshan, E. C. G., and Misra, B. (1977). Time evolution of unstable quantum states and a resolution of Zeno's paradox. *Phys. Rev. D.* **16**: 520–529.

Chwistek, L. (1932). Die nominalistische Grundlegung der Mathematik. *Erkenntnis* **3**: 367–388.

Chwistek, L. (1949). *The Limits of Science.* London: Routledge & Kegan Paul.

Cini, M. (1983). Quantum theory of measurement without wave packet collapse. *Nuovo Cimento* **73B**: 27–56.

Clauser, J. F. and A. Shimony (1978). Bell's theorem: experimental tests and implications. *Rep. Progr. Phys.* **41**: 1881–1927.

Claverie, P. and S. Diner (1977). Stochastic electrodynamics and quantum theory. *Intern. J. Quantum Chemistry* XII Suppl. 1: 41–82.

Copeland, B. J. (1980). The trouble Anderson and Belnap have with relevance. *Phil. Studies* **37**: 325–334.

Cox, D. R. (1962). *Renewal Theory.* London: Methuen.

Curry, H. B. (1951). *Outlines of a Formalist Philosophy of Mathematics.* Amsterdam: North-Holland.

Da Costa, N. C. A. (1980). *Ensaio sobre os fundamentos da lógica.* São Paulo: HUCITEC–EDUSP.

da Costa, N. C. A. (1982a). Statement of purpose. *J. Non-Classical Logic* **1**: i-v.

Da Costa, N. C. A. (1982b). The philosophical import of paraconsistent logic. *J. of Non-Classical Logic* **1**: 1–19.

Davies, P. (1983). *God and the New Physics.* London: Dent.

Davis, C. (1974). Materialist mathematics. In R. S. Cohen, J. Stachel, and M. Wartofsky, Eds. *For Dirk Struik*, pp. 37–66. Boston: Reidel.

Davis, P. J. and R. Hersh (1981). *The Mathematical Experience.* Boston: Birkhäuser.

de Finetti, B. (1972). *Probability, Induction and Statistics.* New York: John Wiley.

de la Peña-Auerbach, L. (1969). New formulation of stochastic theory and quantum mechanics. *J. Math. Phys.* **10**: 1620–1630.

de la Peña, L. (1979). *Introducción a la mecánica cuántica.* México: Compañía Editora Continental.

d'Espagnat, B. (1973). Quantum logic and non-separability. In J. Mehra, Ed., *The Physicist's Conception of Nature* pp. 714–735. Dordrecht: Reidel.

d'Espagnat, B. (1979). The quantum theory and reality. *Sci. Amer.* **241**, No. 5: 158–181.

d'Espagnat and M. Paty (1980). La physique et le réel. *Bull. Soc. franç. Philo.*, Vol. 74, No. 1: 1–42.

deWitt, B. S. and N. Graham, Eds. (1973). *The Many-Worlds Interpretation of Quantum Mechanics.* Princeton: Princeton University Press.

Dieudonné, J. (1981). Bourbaki et la philosophie des mathématiques. *Epistemologia* **4**: 173–188.

Dirac, P. A. M. (1958). *The Principles of Quantum Mechanics*, 4th ed. Oxford: Clarendon Press.

Dirac, P. A. M. (1982). Pretty mathematics. *Int. J. Theor. Phys.* **21**: 603–605.

Dummett, M. (1977). *Elements of Intuitionism.* Oxford: Clarendon Press.

Eberhard, P.H. (1982). Constraints of determinism and of Bell's inequalities are not equivalent. *Phys. Rev. Lett.* **49**: 1474–1477.

Edmonds Jr, J. D. (1978). The muon clock, time dilation and the dynamic vacuum. *Speculations in Sci. and Tech.* **1**: 21–27.

Einstein, A. (1948). Quantum theory and reality. *Dialectica* **2**: 320–324.

Einstein, A., B. Podolsky and N. Rosen (1935). Can quantum-mechanical description of reality be considered complete? *Phys. Rev.* **47**: 777–780.

Enriques, F. (1913). *Les concepts fondamentaux de la science.* Paris: Flammarion.

Enriques, F. (1940). *Problems of Science.* La Salle: Open Court.

Everett III, H. (1957). "Relative state" formulation of Quantum Mechanics. *Revs. Mod. Physics* **29**: 454–462.

Everett III, H. (1973). The theory of the universal wave function. In B. S. DeWitt and N. Graham, Eds., pp. 3–140.

Feynman, R.P., R. Leighton and M. Sands (1963). *The Feynman Lectures on Physics*, Vol. I. Reading, Ma.: Addison-Wesley.

Field, G. B., H. Arp, and J. N. Bahcall (1973). *The Redshift Controversy.* Reading, Mass.: W. A. Benjamin.

Field, H. (1980). *Science Without Numbers.* Princeton: Princeton University Press.

Fine, A. (1982). Hidden variables, joint probabilities and the Bell inequalities. *Phys. Rev. Letters* **48**: 291–295.

Fine, T. (1973). *Theories of Probability: An Examination of Foundations.* New York: Academic Press.

Fonda, L., G. C. Ghirardi, and A. Rimini, (1978). Decay theory of unstable quantum systems. *Rep. Progr. Phys.* **41**: 587–631.

Fourman, M. (1977). The logic of topoi. In Barwise Ed. pp. 1053–1090.

Fox, J. R. (1983). Nonreduction of the state vector in quantum measurement. *Am. J. Phys.* **51**: 49–53.

Fraenkel, A. A. (1966). *Set Theory and Logic.* Reading, Mass.: Addison-Wesley.

Fraïssé, R. (1982). Les mathématiques ne sont-elles qu'un jeu? In R. Apéry et al., *Penser les mathématiques*, pp. 39–57. Paris: Ed. du Seuil.

Frank, P. (1938). *Interpretations and Misinterpretations of Modern Physics.* Paris: Hermann.

Frank, P. (1946). *Foundations of Physics.* Chicago: University of Chicago Press.

Fréchet, M. (1946). Les définitions courantes de la probabilité. In *Les mathématiques et le concret*, pp. 157–204. Paris: Presses Universitaires de France, 1955.

Fréchet, M., Ed. (1967). *Emile Borel philosophe et homme d'action.* Paris: Gauthier-Villars.

Frege, G. (1884). *The Foundations of Arithmetic.* Transl. J. L. Austin, 2nd ed. Oxford: Basil Blackwell & Mott.

Frege, G. (1885). Über formale Theorien der Arithmetik. In *Kleine Schriften* pp. 103–111. I. Angelelli, Ed. Hildesheim: Georg Olms, 1967.

Friedman, M. (1983). *Foundations of Space-Time Theories.* Princeton, N.J.: Princeton University Press.

Friedman, M. and H. Putnam (1978). Quantum logic, conditional probability, and interference. *Dialectica* **32**: 305–315.

Gabbay, D.M. and F. Guenthner, Eds. (1983-85) *Handbook of Philosophical Logic*, 4 vols. Dordrecht-Boston: Reidel.

García-Sucre, M. and M. Bunge (1981). Geometry of a quantal system. *Int. J. Quantum Chem*, **19**: 83–93.

Giles, R. (1964). *Mathematical Foundations of Thermodynamics.* Oxford: Pergamon.

Glansdorff, P. and I. Prigogine (1971). *Thermodynamic Theory of Structure, Stability and Fluctuations.* New York: Wiley-Interscience.

Gödel, K. (1947). What is Cantor's continuum problem? *Amer. Math. Monthly* **54**: 515–525.

Gödel, K. (1951). Russell's Mathematical Logic. In P. A. Schilpp, Ed., *The Philosophy of Bertrand Russell* pp. 125–153.

Goldblatt, R. (1979). *Topoi: The Categorial Analysis of Logic.* Amsterdam: North-Holland.

Gombrich, E. (1961). *Art and Illusion*, 2nd ed. Princeton University Press.

Goodman, N. (1956). A world of individuals. In *The Problem of Universals.* Notre Dame; Notre Dame University Press.

Goodman, N. D. (1979). Mathematics as an objective science. *Am. Math. Monthly* **86**: 540–555.

Goodman, N. D. (1983). Reflections on Bishop's philosophy of mathematics. *Math. Intelligencer* **5**, No. 3: 61–68.

Goodstein, R. L. (1969). Empiricism in mathematics. *Dialectica* **23**: 50–57.

Gottfried, K. (1966). *Quantum Mechanics*, Vol. I. New York: W. A. Benjamin.

Grassmann, H. (1844). *Ausdehnungslehre.* In *Gesammelte mathematische und physikalische Werke*, Vol. I, Part 1. Leipzig, 1894. Repr.: Bronx, N.Y.: Chelsea, 1969.

Grünbaum, A. (1973). *Philosophical Problems of Space and Time*, rev. ed. Dordrecht: Reidel.

Grzegorczyk, A. (1959). Some approaches to constructive analysis. In A. Heyting, Ed. *Constructivity in Mathematics.* Amsterdam: North-Holland.

Haack, S. (1978). *Philosophy of Logics.* Cambridge: Cambridge University Press.

Hacking, I. (1979). What is logic? *J. Philosophy* **76**: 285–319.

Hall, J., C. Kim, B. McElroy, and A. Shimony (1977). Wave-packet reduction as a medium of communication. *Founds. Phys.* **7**: 759–767.

Hammond, A. A. (1978). Mathematics – our invisible culture. In Steen Ed. pp. 15–34.

Hardy, G. H. (1967). *A Mathematician's Apology* London: Cambridge University Press.

Harper, W. L. and C. A. Hooker, Eds. (1976) *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, 3 vols. Dordrecht: Reidel.

Harrison, E. R. (1981). *Cosmology.* Cambridge: Cambridge University Press.

Harsanyi, J. C. (1983). Mathematics, the empirical facts, and logical necessity. *Erkenntnis* **19**: 167–192.

Hartnett, W. E. (1963). *Principles of Modern Mathematics*, 2 vols. Glenview, Ill.: Scott, Foresman.

Hatcher, W. (1982). *The Logical Foundations of Mathematics.* Oxford-New York: Pergamon.

Heisenberg, W. (1930). *The Physical Principles of the Quantum Theory.* Chicago: University of Chicago Press.

Heisenberg, W. (1958). *Physics and Philosophy* New York: Harper & Brothers.

Heisenberg, W. (1969). *Der Teil und das Ganze.* München: Piper.

Hellman, R. H. G., Ed. (1980). *Nonlinear Dynamics. Annals New York Acad. Sci.*, v. 357.

Hepp, K. (1972). Quantum theory of measurement and macroscopic observables. *Helvetica Physica Acta* **45**: 237–248.

Heyting, A. (1956a). La conception intuitionniste de la logique. *Les études philosophiques* **11**: 226–233.

Heyting, A. (1956b). *Intuitionism: An Introduction.* Amsterdam: North-Holland.

Hilbert, D. (1925). Ueber das Unendliche. *Math. Annalen* **95**: 161–190.

Hilbert, D. (1935). *Gesammelte Abhandlungen*, Vol. 3. Berlin: Springer.

Hilbert, D. and P. Bernays (1968, 1970). *Grundlagen der Mathematik*, 2nd ed., 2 vols. Berlin-Heidelberg-New York: Springer-Verlag.

Hintikka, J., Ed. (1969). *The Philosophy of Mathematics.* Oxford: Oxford University Press.

Holton, G. (1973). *Thematic Origins of Scientific Thought.* Cambridge, Ma.: Harvard Univ. Press.

Horwitz, L. P. and E. Katznelson (1983). Is proton decay measurable? *Phys. Rev. Let.* **50**: 1184–1186.

Hoyle, F. and G. (1963). *The Fifth Planet.* New York: Harper & Row.

Hughes, G. E. and M. J. Cresswell (1968). *An Introduction to Modal Logic.* London: Methuen.

Jaki, S. (1974). *Science and Creation.* New York: Science History Publications.

Jammer, M. (1974). *The Philosophy of Quantum Mechanics.* New York: John Wiley & Sons.

Jastrow, R. (1978). *God and the Astronomers.* New York: Norton.

Jaynes, E. T. (1979). Where do we stand on maximum entropy? In R. Levine and M. Tribus, Eds., *The Maximum Entropy Formalism* pp. 14–118. Cambridge, MA: MIT Press.

Jech, T. J. (1973). *The Axiom of Choice.* Amsterdam: North-Holland.

Jeffreys, H. (1957). *Scientific Inference*, 2nd ed. Cambridge: Cambridge University Press.

Johnston, P. T. (1977). *Topos Theory.* London: Academic Press.

Jönsson, C. (1974). Electron diffraction at multiple slits. *Am. J. Physics* **42**: 4–11.

Kálmar, L. (1967). Foundations of mathematics – whither now? In Lakatos, Ed. pp. 187–194.

Kálnay, A. J. (1971). The localization problem. In Bunge Ed. 1971 pp. 93–110.

Kanitscheider, B. (1979). *Philosophie und Moderne Physik.* Darmstadt: Wissenschaftliche Buchgesellschaft.

Kemble, E. C. (1937). *The Fundamental Principles of Quantum Mechanics.* New York: McGraw Hill.

Kitcher, P. (1983). *The Nature of Mathematical Knowledge.* New York: Oxford University Press.

Kitts, D. B. (1974). Physical theory and geological knowledge. *J. Geol.* **82**: 1–23.

Kitts, D. B. (1977). *The Structure of Geology.* Dallas: Southern Methodist University Press.

Kleene, S. C. (1952). *Introduction to Metamathematics.* Princeton N.J.: Van Nostrand.

Kneale, W. C. (1972). Numbers and numerals. *Brit. J. Phil. Sci.* **23**: 191–206.

Kneale, W. C. and M. (1962). *The Development of Logic.* Oxford: Clarendon Press.

Koyré, A. (1943). Galileo and Plato. *Journal of the History of Ideas* **4**: 400–428.

Kretschmann, E. (1917). Über den physikalischen Sinn der Relativitätspostulate. *Annalen der Physik* **53**: 575–614.

Krüger, L. (1976). Reduction versus elimination of theories. *Erkenntnis* **10**: 295–309.

Kuyk, W. (1977). *Complementarity in Mathematics.* Dordrecht-Boston: Reidel.

Lakatos, I. (1976). *Proofs and Refutations.* J. Worrall and E. Zahar, Eds. Cambridge: Cambridge University Press.

Lakatos, I. (1978). *Philosophical Papers*, 2 volumes. Cambridge: Cambridge University Press.

Lakatos, I., Ed. (1967). *Problems in the Philosophy of Mathematics.* Amsterdam: North-Holland.

Lalande, A., Ed. (1938). *Vocabulaire technique et critique de la philosophie*, 4th ed., 3 vols. Paris: Alcan.

Lambek, J. (1982). The influence of Heraclitus on modern mathematics. In Agassi and Cohen, Eds., pp. 111–121.

Landau, L. D. and E. M. Lifshitz (1958). *Quantum Mechanics. The Non-Relativistic Theory.* Reading, Mass.: Addison-Wesley.

Lawvere, F. W. (1966). The category of categories as a foundation for mathematics *Proc. La Jolla Conference on Categorical Algebra* pp. 1–20. New York: Springer.

Leibniz, G. W. (1703). *Nouveaux essais.* Engl. transl. P. Remnant and J. Bennett, *New Essays on Human Understanding.* Cambridge: Cambridge University Press, 1981.

Leontief, W. (1982). Academic economics. *Science* **217**: 104–107.

Leśniewski, S. (1927-31). On the foundations of mathematics. Transl. V. F. Sinisi. *Topoi* **2**: 7–52 (1983).

Levi-Civita, T. (1929). *The Absolute Differential Calculus.* London and Glasgow: Blackie & Son.

Levine, R. D. and R. B. Bernstein (1974). *Molecular Reaction Dynamics.* Oxford: Clarendon Press.

Lévy, M. (1979). Relations entre chimie et physique. *Epistemologia* II: 337–370.

Lévy-Leblond, J.-M. (1976). One more derivation of the Lorentz transformation. *Am. J. Phys.* **44**: 271–277.

Lévy-Leblond, J.-M. (1977). Towards a proper quantum theory. In J. Leite Lopes and M. Paty, Eds., *Quantum Mechanics, A Half Century Later*, pp. 171–206. Dordrecht-Boston: Reidel.

Lévy-Leblond, J.-M. and F. Balibar (1984). *Quantique.* Paris: InterEditions.

Lewis, C. I. and C. H. Langford (1959). *Symbolic Logic*, 2nd ed. New York: Dover.

Liang, E. P. T., and R. K. Sachs (1980). Cosmology. In A. Held, Ed. *General Relativity and Gravitation.* New York: Plenum Press.

London, F. and E. Bauer (1939). *La théorie de l'observation en mécanique quantique.* Paris: Hermann.

Łukasiewicz, J. (1970). *Selected Works.* L. Borkowski Ed. Amsterdam: North-Holland.

Machida, S. and Namiki, M. (1980). Theory of measurement in quantum mechanics. *Progress of Theor. Phys.* **63**: 1457–1473; 1833–1847.

MacLane, S. (1971). *Categories for the Working Mathematician.* New York: Springer-Verlag.

MacLane, S. (1981). Mathematical models: a sketch for the philosophy of mathematics. *Amer. Math. Monthly* **88**: 462–472.

Margenau, H. (1936). Quantum-mechanical description. *Phys. Rev.* **49**: 240–242.

Markow, A. A. (1975). *Was ist konstruktive Logik?* Berlin: Urania.

Marshak, J. *et al.* (1975). Personal probabilities of probabilities. *Theory and Decision* **6**: 121–159.

Martin-Löf, A. (1979). Statistical mechanics and the foundations of thermodynamics. *Lecture Notes in Physics* No. 101.

Martino, E. (1982). Creative subject and bar theorem. In Troelstra and Van Dalen, Eds., pp. 311–318.

Mill, J. S. (1843). *A System of Logic*. London: Longmans, Green, 1952.

Miller, W. H. *et al.*, Eds. (1976). *Modern Theoretical Chemistry*, Vols. 1 and 2. New York: Plenum Press.

Miró-Quesada, F. (1983). Paraconsistent logic: some philosophical issues. Preprint.

Misra, B. and Sudarshan, E. C. G. (1977). The Zeno's paradox in quantum theory. *J. Math. Phys.* **18**: 756–763.

Moore, G. H. (1982). *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*. New York: Springer-Verlag.

Mosterín, J. (1978). *Racionalidad y acción humana*. Madrid: Alianza Editorial.

Muncaster, R. G. and C. Truesdell (1979). *Fundamentals of Maxwell's Kinetic Theory*. New York: Academic Press.

Munitz, M. K., Ed. (1957). *Theories of the Universe*. Glencoe, Ill.: Free Press.

Newton, R. G. (1980). Probability interpretation of quantum mechanics. *Am. J. Phys.* **48**: 1029–1034.

Neyman, J. (1963). Stochastic approach to cosmology. In S. Drobot, Ed., *Mathematical Models in Physical Sciences*. Englewood Cliffs, N. J.: Prentice-Hall.

Noll, W. (1967). Space-time structures in classical mechanics. In Bunge Ed. 1967a pp. 28–34.

Nordin, I. (1979). Determinism and locality in quantum mechanics. *Synthèse* **42**: 71–90.

Paty, M. (1981). L'inséparabilité quantique en perspective. Centre de Recherches Nucléaires de Strasbourg HE 81-10.

Pauli, W. (1958). Die allgemeinen Principien der Wellenmechanik. In S. Flügge, Ed., *Handbuch der Physik*, Vol. V, Teil 1, pp. 1–168. Berlin: Springer-Verlag.

Pauli, W. (1961). *Aufsätze und Vorträge über Physik und Erkenntnistheorie*. Braunschweig: Vieweg & Sohn.

Perdang, J. (1983). On physical proofs of mathematical theorems. *Phys. Letters* **93A**: 459–463.

Peres, A. (1980). Zeno paradox in quantum theory. *Am. J. Phys.* **48**: 931–932.

Petersen, A. (1963). The philosophy of Niels Bohr. *Bull. Atomic Sci.* 19, No. 7: 8–14.

Poincaré, H. (1903). *La science et l'hypothèse*. Paris: Flammarion.

Polanyi, J. C. and J. L. Schreiber (1973). The dynamics of bimolecular collisions. In H. Eyring, J. Henderson, and W. Jost, Eds., *Physical Chemistry: An Advanced Treatise*, Vol. VIA, pp. 383–487. New York: Academic Press.

Pólya, G. (1954). *Mathematics and Plausible Reasoning*, 2 volumes. Princeton, N.J: Princeton University Press.

Popper, K. R. (1957b). The propensity interpretation of the calculus of probability and the quantum theory. In S. Körner, Ed., *Observation and Interpretation*, pp. 65–70. London: Butterworths Scientific Publ.

Post, E. J. (1977). A minor and a major predicament of physical theory. *Foundations of Physics* 7: 255–277.

Prélat, C. E. (1947). *Epistemología de la quimica*. Buenos Aires: Espasa-Calpe.

Primas, H. (1983). *Chemistry, Quantum Mechanics and Reductionism.* New York: Springer-Verlag.

Prior, A. N. (1967). *Past, Present and Future.* Oxford: Oxford University Press.

Putnam, H. (1975). *Philosophical Papers*, 2 volumes. Cambridge University Press.

Putnam, H. (1976). How to think quantum logically. In P. Suppes, Ed. *Logic and Probability in Quantum Mechanics.* Dordrecht-Boston: Reidel.

Quine, W. v. O. (1969). *Ontological Relativity and Other Essays.* New York: Columbia University Press.

Quine, W. v. O. (1970). *Philosophy of Logic.* Englewoods Cliffs, N.J.: Prentice-Hall.

Ramsey, F. P. (1931). *The Foundations of Mathematics.* London: Routledge & Kegan Paul.

Ramsey, N. F. (1956). Thermodynamics and statistical mechanics at negative absolute temperatures *Phys. Rev.* **103**: 20–28.

Rasiowa, H. (1974). *An Algebraic Approach to Non-Classical Logics.* Amsterdam: North-Holland.

Rasiowa, H. and R. Sikorski (1970). *The Mathematics of Metamathematics* 3rd ed. Warsaw: PWN.

Reichenbach, H. (1949). *The Theory of Probability.* Berkeley and Los Angeles: University of California Press.

Rescher, N. (1969). *Many-Valued Logic.* New York: McGraw-Hill.

Rescher, N. (1977a). *Methodological Pragmatism.* New York: New York University Press.

Rescher, N. and A. Urquhart (1971). *Temporal Logic.* Wien: Springer.

Restivo, S. (1983). *The Social Relations of Physics, Mysticism, and Mathematics.* Boston: Reidel.

Riccardi, A. C. (1977). Geología: protociencia, especulación o ciencia? *Rev. Asoc. Geol. Argentina* **32**: 52–69.

Robinson, A. (1965). Formalism 64. In Y. Bar-Hillel, Ed., *Logic, Methodology and Philosophy of Science*, pp. 228–246. Amsterdam: North-Holland.

Robinson, J. (1962). *Economic Philosophy.* Hardsmonworth: Penguin, 1964.

Rohrlich, F. (1983). Facing quantum mechanical reality. *Science* **221**: 1251–1255.

Rosenfeld, L. (1953). Strife about complementarity. *Science Progress* No. 163: 393–410, being a rev. version of "L'évidence de la complémentarité", in A. George, Ed., *Louis de Broglie, physicien et penseur* Paris: Albin Michel, 1953.

Rosenfeld L. (1961). Foundations of quantum theory and complementarity. *Nature* **190**: 384–388.

Rosenfeld, L. (1964). Discussion. In L. Infeld, Ed. *Conférence internationale sur les théories relativistes de la gravitation* pp. 219–220. Paris: Gauhier-Villars; Warsaw: PWN.

Routley, R. (1980). *Exploring Meinong's Jungle and Beyond*, interim edition. Canberra: Australian National University.

Ruelle, D. (1969). *Statistical Mechanics. Rigorous Results.* New York: Benjamin.

Ruelle, D. (1978). *Thermodynamic Formalism.* Reading, Ma.: Addison-Wesley.

Russell, B. (1901). Mathematics and the metaphysicians. In *Mysticism and Logic.* London: Pelican, 1953.

Rutherford, D. E. (1965). *Introduction to Lattice Theory.* Edinburgh: Oliver & Boyd.

Savage, L. J. (1972). *The Foundations of Statistics*, 2nd ed. New York: Dover.

Schiff, L. I. (1949). *Quantum Mechanics* New York: McGraw-Hill.

Schilpp, P.A., Ed. (1949). *Albert Einstein: Philosopher-Scientist.* Evanston, Ill.: The Library if Living Philosophers.

Schrödinger, E. (1935a). Die gegenwärtige Situation in der Quantenmechanik. *Naturwissenschaften* **23**: 807–812, 823–828, 844–849.

Schrödinger, E. (1935b). Discussion of probability relations between separated systems. *Proc. Cambridge Phil. Soc.* **31**: 555–563.

Selleri, F. and Tarozzi, G. (1978). Is nondistributivity for microsystems empirically founded? *Nuovo Cimento* **43**: 31–40.

Settle, T. (1974). Induction and probability unfused. In P. A. Schilpp, Ed. *The Philosophy of Karl Popper* pp. 697–749. La Salle, Ill.: Open Court.

Shankland, R. S. (1963). Conversations with Einstein. *Amer. J. Physics* **31**: 47–57. Conclusion: Ibid. **31**: 895–901.

Sierpiński, W. (1958). *Cardinal and Ordinal Numbers.* Warsaw: Polska Akademia Nauk.

Simpson, G. G. (1963). Historical science. In Albritton Ed. pp. 24–48.

Slater, J. C. (1929). Physical meaning of wave mechanics. *J. Franklin Inst.* **207**: 449–455.

Smith, D. G., Ed. (1981). *The Cambridge Encyclopedia of Earth Sciences.* New York: Crown Publishers.

Smoluchowski, M. von (1918). Über den Begriff des Zufalls und den Ursprung der Wahrscheinlichkeitsgesetze in der Physik. *Naturwissenschaften* **6**: 253–263.

Smorynski, C. (1983). Mathematics as a cultural system. *The Mathematical Intelligencer* **5**: 9–15.

Sneed, J. (1971). *The Structure of Mathematical Physics.* Dordrecht: Reidel.

Stapp, H. (1980). Locality and reality. *Founds. Phys.* 10: 767–795.

Steen, L. A., Ed. (1978). *Mathematics Today.* New York: Springer-Verlag.

Stegmüller, W. (1976). *The Structure and Dynamics of Theories.* New York: Springer-Verlag.

Steiner, M. (1975). *Mathematical Knowledge.* Ithaca, N.Y.: Cornell University Press.

Stoll, R. R. (1963). *Set Theory and Logic.* San Francisco: W. H. Freeman.

Stolzenberg, G. (1970). Review of Bishop's book (1967). *Bull. Amer. Math. Soc.* **76**: 301–323.

Tarski, A. (1956). *Logic, Semantics, Metamathematics.* Oxford: Clarendon Press.

Tarski, A. (1965). A simplified formalization of predicate logic with identity. *Archiv für mathematische Logik und Grundlagenforschung* **7**: 61–79.

Tarski, A., A. Mostowski, and R. M. Robinson (1968). *Undecidable Theories.* Amsterdam: North-Holland.

Theobald, D. W. (1976). Some considerations on the philosophy of chemistry. *Chem. Soc. Rev.* (London) **5**: 203–213.

Thom, R. (1972). *Stabilité structurelle et morphogenèse.* Reading, MA.: W. A. Benjamin.

Tolman, R. C. (1949). The age of the universe. *Rev. Mod. Phys.* **21**: 374.

Torretti, R. (1978). *Philosophy of Geometry from Riemann to Poincaré.* Boston: Reidel.

Torretti, R. (1982). Three kinds of mathematical fictionalism. In Agassi and Cohen Eds. pp. 399–414.

Torretti, R. (1983). *Relativity and Geometry.* Oxford: Pergamon.

Troelstra, A. S. and D. van Dalen, Eds. (1982). *The L. E. J. Brouwer Centenary Symposium.* Amsterdam: North-Holland.

Truesdell, C. (1974). A simple example of an initial-value problem, etc. *Rendiconti Istituto Lombardo di Scienze e Lettere* **108**: 301–304.

Truesdell, C. (1982a). Our debt to the French tradition: "catastrophes" and our search for structure today. *Scientia* **76**: 63–77.

Truesdell, C. (1982b). The disastrous effects of experiment upon the early development of thermodynamics. In Agassi and Cohen Eds. pp. 415–423.

Truesdell, C. (1984). *An Idiot's Fugitive Essays on Science.* New York: Springer-Verlag.

Van Fraassen, B. (1982). The Charybdis of realism: epistemological implications of Bell's inequality. *Synthèse* **52**: 25–38.

Venn, J. (1888). *The Logic of Chance*, 3rd ed. London; Macmillan.

Ville, J. (1939). *Etude critique de la notion de collectif.* Paris: Gauthier-Villars.

von Mises, R. (1972). *Wahrscheinlichkeit, Statistik und Wahrheit*, 4th ed. Wien: Springer-Verlag.

von Neumann, J. (1932). *Die mathematische Grundlagen der Quantenmechanik.* Berlin: Springer. Engl. transl.: *Mathematical Foundations of Quantum Mechanics.* Princeton: Princeton University Press, 1955.

Wald, A. (1950). *Statistical Decision Functions.* New York: John Wiley.

Wang, H. (1966). Process and existence in mathematics. In Y. Bar-Hillel *et al.*, Eds., *Essays on the Foundations of Mathematics. Dedicated to A. A. Fraenkel* pp. 328–351. Jerusalem: Magnus Press.

Wang, H. (1974). *From Mathematics to Philosophy.* London: Routledge & Kegan Paul.

Wedberg, A. (1955). *Plato's Philosophy of Mathematics.* Stockholm: Almqvist & Wiksell.

Weyl, H. (1949). *Philosophy of Mathematics and Natural Science*, rev. ed. Princeton: Princeton University Press.

Weyl, H. (1950). *Space-Time-Matter*, 4th ed. New York Dover.

Wheeler, J. A. (1978). Not consciousness but the distinction between the probe and the probed as central to the elemental act of observation. In Jahn, R. G., Ed. *The Role of Consciousness in the Physical World.* pp. 87–111. Boulder, Colo.: Westview Press.

Wheeler, J. A. and W. H. Zurek, Eds. (1983). *Quantum Theory and Measurement.* Princeton: Princeton University Press.

Whewell, W. (1847). *The Philosophy of the Inductive Sciences*, 2 vols. London: Frank Cass, 1967.

Whitehead, A. N. (1898). *A Treatise on Universal Algebra* I. Cambridge: University Press.

Wigner, E. P. (1960). The unreasonable effectiveness of mathematics in the natural sciences. *Comm. Pure & Applied Math.* **13**: 1–14.

Wigner, E. P. (1963). The problem of measurement. *Am. J. Phys.* **31**: 6–15.

Wilder, R. L. (1952). *An Introduction to the Foundations of Mathematics.* New York: Wiley.

Wilder, R. L. (1981). *Mathematics as a Cultural System.* Oxford and New York: Pergamon.

Wittgenstein, L. (1978). *Remarks on the Foundations of Mathematics*, rev. ed. Oxford: Blackwell.

Woolley, R. G. (1978). Must a molecule have a shape? *J. Am. Chem. Soc.* **100**: 1073–1078.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control* **8**: 338–353.

Zadeh, L. A. (1975). Fuzzy logic and approximate reasoning. *Synthèse* **30**: 407–428.

Zeeman, E. C. (1964). Causality implies the Lorentz group. *J. Math. Phys.* **5**: 490–493.

Zinov'ev, A. A. (1973). *Foundations of the Logical Theory of Scientific Knowledge* (*Complex Logic*). Dordrecht: Reidel.

NAME INDEX

Albritton, C. 231
Alexander 37
Alfvén, H. 241
Anderson, A. R. 56, 63
Angel, R. 165
Apollonius 15
Arbib, M. 79
Archimedes 15, 19, 36, 83, 84
Aris, R. 221
Aristotle 2, 3, 37, 58, 62, 63, 167
Arruda, A. 65
Aspect, A. 169, 205, 210, 211
Axelrad, D. 147

Bachelard, G. 220
Bacon, F. 2
Balibar, F. 186
Ballentine, L. 180
Bar Hillel, M. 92
Barwise, J. 17, 40
Bauer, E. 133, 196
Bell, J. S. 119, 202, 208, 209, 210, 211, 213, 214, 215, 216
Bellman, R. 67
Beinap, N. 56, 63
Benacerraf, P. 108
Bergson, H. 2, 35, 158
Berkeley, G. 2, 170, 175
Bernays, P. 17, 19, 95, 100, 110, 113
Bernstein, R. 231
Beth, E. 95, 98
Birkhoff, G. 67
Bishop, E. 21, 97, 102, 103, 104, 105, 115, 116
Blokhinzev, D. 180
Bôcher, M. 62
Bohm, D. 167, 168, 184, 195, 206, 213
Bohr, N. 133, 167, 168, 169, 171, 174, 175, 185, 196
Boltzmann, L. 151, 153
Bolzano, B. 2, 12, 23
Borel, E. 117
Born, M. 166, 178, 179, 180, 192
Bourbaki, N. 19, 23, 55, 101, 112
Bridgman, P. 161
Broad, C. D. 220
Brouwer, L. E. J. 19, 21, 28, 54, 55, 61, 70, 98, 100, 102, 103, 115, 116
Browder, F. 85
Brunschwicg, L. 108
Brush, S. 153
Bunge, M. 11, 26, 27, 28, 41, 46, 52, 59, 66, 70, 80, 87, 89, 93, 94, 97, 100, 101, 115, 117, 119, 129, 135, 138, 144, 145, 148, 153, 155, 157, 159, 161, 165, 167, 168, 169, 170, 171, 175, 183, 190, 196, 203, 105, 207, 216, 217, 220, 225, 226, 229, 237, 240
Burbidge, G. 239, 240

Caldin, E. 220
Cantor, G. 86
Carnap, R. 90
Carr, B. 241
Carter, B. 241
Casanova, G. 99, 117
Cauchy, A. 24, 25
Chiu, C. 203
Chwistek, L. 112
Cini, M. 196, 202
Clauser, J. F. 209, 210, 211
Claverie, P. 206
Copeland, B. 64
Cox, D. 85
Cresswell, M. 63
Curry, H. 100, 102

d'Espagnat, B. 68, 168, 210, 211, 215

da Costa, N. 56, 65
Davies, P. 239
Davis, P. 18, 100, 117
de Broglie, L. 168, 184, 206
de Finetti, B. 90
de la Peña-Auerbach, L. 180, 206
de Witt, B. 201
Descartes, R. 2
Dewey, J. 2
Diderot, D. 2
Dieudonné, J. 101, 112
Diner, S. 206
Dirac, P. A. M. 42, 86, 136, 189, 198
Doppler, C. 145, 239
Dummett, M. 28, 54, 97, 98, 115

Eberhard, P. 215
Edmonds, J. 160
Einstein, A. 86, 157, 160, 168, 175, 180, 184, 205, 206, 207, 208, 210, 211, 212, 213, 214, 215, 216, 217, 236, 238, 239, 240
Enriques, F. 117
Euclid 15, 23, 30, 38, 142
Eudoxus 15
Everett, H. 196, 200, 201, 203

Fermi, E. 134
Feynman, R. P. 89, 169, 217
Field, G. B. 239
Field, H. 27, 112
Fine, A. 87, 215
Fine, T. ?
Fonda, L. 193
Fourier, J.-B.-J. 154
Fourman, M. 30
Fox, J. R. 201
Fraenkel, A. 51
Fraïssé, R. 101, 113
Frank, P. 168
Fréchet, M. 26, 89, 93, 117
Frege, G. 2, 21, 23, 42, 58, 98, 111, 113
Friedman, M. 68, 165, 190, 207, 238

Gabbay, D. 55
Galilei, G. 139, 144, 156, 157, 167, 181
García-Sucre, M. 225
Gauss, C. F. 142
Gibbs, J. W. 151, 180
Giles, R. 154
Glansdorff, P. 152
Gödel, K. 17, 40, 49, 50, 101, 105, 111
Goldblatt, R. 96
Gombrich, E. 84
Gonseth, F. 13
Good, I. J. 90
Goodman, N. 102, 111, 114, 119
Goodstein, R. 119
Gottfried, K. 194, 202
Graham, N. 201
Grassmann, H. 23, 115
Grünbaum, A. 135, 136, 160, 162, 165
Grzegorczyk, A. 105
Guenther, F. 55

Haack, S. 56
Hacking, I. 53
Hall, J. 195
Hallett, M. 119
Hammond, A. 122
Hardy, G. 115
Harper, W. 93
Harrison, E. 236
Harsanyi, J. 117
Hartnett, W. 19
Hatcher, W. 95, 98
Hegel, G. W. F. 2, 29, 235
Heidegger, M. 2
Heisenberg, W. 49, 168, 169, 171, 181, 182, 183, 184, 186, 187, 189, 200, 218
Hellman, R. 148
Hepp, K. 196, 202
Hermite, C. 111, 174
Hersch, R. 18, 100
Heyting, A. 58, 59, 60, 102, 115
Hilbert, D. 17, 49, 95, 99, 100, 101, 113, 173, 174, 199
Hiley, B. 213
Hintikka, J. 17, 90
Hobbes, T. 2
Holton, G. 157
Hooker, C. 93
Horwitz, L. 193, 203
Hoyle, F. 158
Hoyle, G. 158
Hubble, E. 236

Hughes, G. 63
Hume, D. 2
Husserl, E. 2
Huyghens, C. 187, 189
Hypatia of Alexandria 16

Jaki, S. 239
Jammer, M. 167
Jastrow, R. 238, 239
Jauch, J. M. 67
Jaynes, E. 150
Jech, T. 51
Jeffreys, H. 90
Johnston, P. 96
Jönsson, C. 187

Kálmar, L. 12, 13, 117
Kálnay, A. J. 128, 196, 203
Kanitscheider, B. 236
Kant, I. 2, 102, 107, 141, 142, 237
Katznelson, E. 193, 203
Kemble, E. 179
Kepler, J. 137
Kitcher, P. 117
Kitts, D. 231, 232, 234
Kleene, S. 17
Klein, F. 48
Kneale, W. 53, 113
Kolmogoroff, A. 87
Koyré, A. 78
Kretschmann, E. 139, 140
Krüger, L. 153
Kuhn, T. 234
Kuyk, W. 117

Lakatos, I. 13, 18, 95, 100, 110, 117
Lalande, A. 111
Lambek, J. 36
Landau, L. 167
Langford, C. H. 63
Laplace, P. S. 37, 183, 237
Lawvere, W. 52
Leibniz, G. W. 2, 11, 12, 83, 123, 217
Lenin, V. 67
Leontief, W. 85
Lesniewski, S. 112
Levi-Civita, T. 165
Levine, R. 231
Lévy, M. 220, 231
Lévy-Leblond, J.-M. 159, 186, 201
Lewis, C. I. 63
Lewis, G. N. 224
Liang, E. 240
Lifshitz, E. 167
Locke, J. 2
London, F. 133, 196
Lorentz, H. A. 139, 144, 156, 157, 158, 159, 161, 213
Lukasiewicz, J. 62

Mach, E. 110, 135, 145, 146
Machida, S. 196, 202
MacLane, S. 17, 19, 52, 85, 118
Margenau, H. 193, 201
Markow, A. 102, 113
Marshak, J. 93
Martin-Löf, A. 153
Martino, E. 115
Marx, K. 2, 3
Maxwell, J. C. 36, 139, 156, 157
Meyerson, E. 220
Michelson, A. A. 157
Mill, J. S. 117
Miller, W. 231
Millikan, R. A. 134
Miró Quesada, F. 58
Misra, B. 203
Moore, G. H. 51, 52
Morley, E. W. 157
Mosterín, J. 73
Muncaster, R. 153
Munitz, M. 235

Nagel, E. 80
Namiki, M. 196, 202
Newton, I. 135, 143, 145, 146, 156, 167, 179, 237, 240
Neyman, J. 240
Nietzsche, F. 123
Noll, W. 143
Nordin, I. 215

Pascal, B. 14, 77
Paty, M. 213
Paty, M. 211
Pauli, W. 166, 168, 169, 194, 224

Peano, G. 20, 101, 115
Perdang, J. 19
Peres, A. 203
Petersen, A. 168
Peirce, C. S. 2
Planck, M. 137, 153, 166, 170, 171
Plato 2, 4, 12, 14, 23, 24, 29, 33, 35, 50, 111, 217, 218, 221
Podolsky, B. 175, 205, 206, 207, 208, 210, 211, 212, 213, 214, 215, 216, 217
Poincaré, H. 55, 89, 116, 135, 136, 157, 160
Polanyi, J. 229
Pólya, G. 13, 18
Popper, K. 2, 4, 89, 90
Post, E. 127
Prélat, C. 220
Prigogine, I. 152
Primas, H. 220, 231
Prior, A. 64
Putnam, H. 13, 68, 108, 117, 118, 190, 207
Pythagoras 241

Quine, W. v. O. 13, 21, 45, 46

Ramsey, F. P. 101, 154
Rasiowa, H. 17, 71
Rees, M. 241
Reichenbach, H. 90, 93, 135
Renyi, A. 87
Rescher, N. 62, 64, 72
Restivo, S. 13, 117
Riccardi, A. 231, 232
Riemann, B. 162
Robertson, H. P. 236, 238
Robinson, A. 31, 100
Robinson, J. ?
Rohrlich, F. 216
Rosen, N. 175, 205, 206, 207, 208, 210, 211, 212, 213, 214, 215, 216, 217
Rosenfeld, L. 133, 168, 196
Routley, R. 27, 65
Ruelle, D. 153
Russell, B. 2, 3, 45, 48, 98, 110
Rutherford, D. 70
Ryle, G. 77

Sachs, R. 240
Savage, L. 90
Schiff, L. 189
Schreiber, J. 229
Schrödinger, E. 170, 171, 172, 175, 176, 177, 178, 179, 189, 198, 199, 200, 202, 204, 212, 213, 214, 215, 225, 226, 229
Selleri, F. 190
Settle, T. 89
Shankland, R. 157
Shimony, A. 195, 209, 210, 211
Sierpinski, W. 51
Sikorski, R. 17
Simpson, G. 232, 233
Slater, J. 180, 218
Smith, D. 234
Smoluchowski, M. 89
Smorynski, C. 17, 49
Sneed, J. 47
Spinoza, B. 2
Stapp, H. 211
Stegmüller, W. 47
Steiner, M. 99
Steno, N. 233
Stieltjes, T. J. 111
Stoll, R. R. 49, 95
Stolzenberg, G. 21, 102, 116
Sudarshan, E. 203
Suppes, P. 101

Tarozzi, G. 190
Tarski, A. 49, 54, 57
Theobald, D. 220
Thom, R. 84
Thurston, W. 111
Torretti, R. 34, 87, 123, 161, 165
Troelstra, A. 102
Truesdell, C. 47, 141, 147, 153, 154
Turing, A. 223

Urquhart, A. 64

Vaihinger, H. 123
van Dalen, D. 102
van Fraassen, B. 215
Venn, J. 93
Ville, J. 93
Voltaire 2
von Mises, R. 93, 98
von Neumann, J. 67, 133, 195, 196, 199, 200, 201, 204

Wald, A. 92
Walker, A. C. 236, 238
Wang, H. 13, 49, 50, 99
Weber, A. 73
Wedberg, A. 12, 111
Wegener, A. 234, 235
Weierstrass, K. 100
Weyl, H. 22, 115, 158
Wheeler, J. A. 132, 192, 216
Whewell, W. 142
Whitehead, A. N. 2, 101
Wigner, E. 84, 132, 196
Wilder, R. 13, 95, 98
Wittgenstein, L. 2, 12, 13, 95, 101, 110
Woolley, R. 225

Zadeh, L. 67, 79
Zeeman, J. 159
Zeno 82, 203, 204, 205
Zermelo, E. 38
Zinov'ev, A. 112
Zurek, W. 192

SUBJECT INDEX

Absoluteness 27
Abstract construct 47–48, 100
Analogy
  classical 171
Analycity 46–47
Applications of mathematics 75–95
  endomathematical 76–77
  exomathematical 76–77
Apriorism 142
Axiomatics 100–102, 113

Background
  formal 10, 14
  philosophical 10. 4
  specific 10, 14
Becoming 158–159
Bell's Theorems 208–213, 216
Big Bang Hypothesis 238–239
Born's Postulate 178–181
Bridges between the humanities and S&T 3–7

Cat paradox 176–177
Category theory 52–53
Cauchy's integral theorem 25
Causality 159–160
Change, mathematization of 83–84
Chemical kinetics 223
Chemistry 219–231
  quantum 223–231
  reduction to physics 225–231
Choice, axiom of 50–52
Classicist interpretation of quantum theory 174–176, 207, 216
Classon 166, 175
Collection, variable 36–37
Commutation equation 184
Complementarity 68, 182
  principle 185–187

Conceivability 28–29
Conceptualism 12
Consciousness 132–133
Construct 11, 29, 33–34
Constructivism 59–60, 97, 103–106
Contradiction 58, 65–67
Controversies over quantum mechanics 167–169
Convention 84, 129, 159
Conventionalism
  mathematical 114–115
  physical 135–136, 160–162
Coordinate transformation 144–145
Copenhagen interpretation 133, 168–169, 172–173, 182–184, 189, 195–197
Correspondence principle 138
Correspondence rule 119
  see also Semantic assumption
Cosmology 235–241
Covariance principles 138–140
CP(= Classical Physics) 166–167
CPT Theorem 138, 139
Creationism in cosmology 238
Criticism 168–169
Curvature of spacetime 163

Dependence of one research field on another 228–229
Determinacy 159
Determinism 183, 206, 207, 211, 215
Dialectics 66
Discovery 23–25
Distributive law 69–70, 190
Dogmatism in physics 170
Double slit experiments 187–191

Earth science 231–235
Eigenvalue 172

Electrodynamics
  classical 156–147
  quantum 166, 167, 185
Embodiments of ideas 111
Emergence 149, 188
Empiricism mathematical 32, 109, 117–119
Entropy 149–150
EPR (Einstein-Podolsky-Rosen)
  correlations 207, 208, 213
  dogma 207, 211
Exactification 78–79
Existence
  conceptual, formal 25–33, 45–46, 115–116
  factual, material, real 29
  statement 29–30

Fallibilism 110
Fiction 38–39, 123
Fictionism
  mathematical 26, 123
Formal/factual dichotomy 10–13, 26, 123
Formalism 99–102, 112–113
Foundational strategy 97–98, 106–107
Frequency, relative 93–95
Friedmann models 238
Fund of knowledge 10, 15
Fuzziness of quantons 176, 184
Fuzzy set 79

Geometry, physical 142–144

Heisenberg's inequalities 181–187
Hidden variables 205–211
Historical discipline 232, 237
Hubble's law 236
Humanities 3

Idealism 23–24
  see also Platonism
Inclusion of one research field in another 227
Incommensurability of theories 167
Incompleteness 49–50
Indeterminacy
  classical 147–148
  quantum 181–187
Indistinguishability 217–218
Individuals 114
Induction 232
Instrumentalist formalism 86
Interpretation 80–82, 87
Intuition 115
Intuitionism 28, 60, 97
  mathematical 59–61, 102–106, 109, 115–116
Invention 23–25
Irreversibility 151, 193
Kant-Laplace hypothesis 237
Knowledge, limits of 186–187

Language 54
Law statement 131, 233
Locality
  see Separability
Logic 40–71
  applied 77–78
  fuzzy 67
  intuitionistic 58–61
  lato sensu 40, 54
  many-valued 62
  modal 63
  non-standard 55–75
  paraconsistent 65–67
  quantum 67–70, 190
  relevance 63–64
  temporal 64–65
Logicism 98–99

Many worlds interpretation of quantum mechanics 201
Mass 145–146
Mathematics 13–123
  applied 22, 75, 95
  foundations of 95–107
  household 110
  philosophy of 107–123
  pure 22
Meaning 20–21, 116
Measurement 131–134, 165, 172–173, 178, 184, 192–200
  obtrusive 194
  theory 133–134, 197
  unobtrusive 194
Mechanics
  classical 140–148

  quantum 165–219
  statistical 148–154
Mensurandum 132
Metamathematics 17
Metanomological statement 138–140
Metatheorem 138–139
Metatheoretical problems 137–140
Method, scientific 15
Methodics 10
Model 47, 79–83
  factual 82
  probabilistic 88
  scientific 82
  technological 82
  theory 47–49
Mutability 27–28

Name 113
Neutrality, epistemological 38
Nominalism 100, 109, 112–115

Objectivity
  methodological 37
  semantical 37
Observer 155, 195–196
Ontology 81
  and logic 46
  and mathematics 34, 37
Operationism 59, 154

Paralogic 56
  see also Logic, paraconsistent, and L, fuzzy
Particle picture 185–186
Philosophy of Science & Technology 1–8
Physical chemistry 223–224
Physics 124–219
Plate tectonics 233–235
Platonism 26, 50, 84, 109, 111–112
Pluralism
  foundational 106
  logical 72
Position coordinate 127–128
Possibility 88, 94
Pragmatism 13
Predicate 41–42
Predication 40–41
Prisoner's Dilemma 93
Probability 86–95
  general concept 87
  logical 90
  objective 88–90, 149–150, 178–180
  personalist, see subjective
  quantum-mechanical 178–187
  subjective 91–93, 150
Problem
  mechanical 146
  metatheoretical 146–148
Problematics 10, 15
Projection of state function 198–205
Property 41
  physical 126–131, 174–175
Proposition 42–44
Protophysics 135

QP (= Quantum Physics) 166–167
Quantifier
  existensial 44–46, 57
Quantity, physical
  see Property, physical
Quanton 166, 171–172, 175–176

Randomness 88, 148, 149
Rationality 66, 73–74
Reaction, chemical 221–222
Realism 175–176, 192, 206, 210–211, 215–216
Realist interpretation of quantum theory 168–205
Reality 34, 83–84
Recursivism 105–106
Redshift 239
Reduction
  of chemistry to physics ?
  on concepts 229
  of mathematics to logic 98–99
  of theories 230
Reference frame 144–145, 155–157
Referent 20–22, 139–140, 157–158, 169–170
  immediate 140
  mediate 140
Relativity
  general theory 161–165
  principle 156–157
  special theory 142, 155–161

Research field 9–10
  factual 10
  formal 10
Reversibility 151
Rigor 23, 72
Robertson-Walker line element 236, 238

S & T (= Science & Technology) 1, 3–4
Satisfaction 47
Schrödinger equation 170–172
Second law of thermodynamics 152–153
Semantic assumption 20, 128–129, 169
Semantics 20
Semiquanton 166, 225
Separability 207, 211–215
Set theory 51–52
Simultaneity 160–161
Sociologism 13
Spacetime 158, 163–165
Spin 175, 208–209
Standard 135–136
State 129–130
State function 170–174, 178–179
State space 173–174
Stern-Gerlach apparatus 133, 195, 198, 202
Structure, mathematical
  see System, conceptual
Subjectivism 126, 149–150, 155
Superposition principle 173–177, 189, 204
Symmetry 218–219
System
  conceptual 19–20
  physical 213–215

T-invariance 147, 151
Temperature 154
Theory 25
  fundamental 136–137
  phenomenological 136
  physical 134–137
  special 137
Thermodynamics 152–154
Thing, physical 125–126
Timelessness of mathematical constructs 35–37
Trend 233
Truth
  of fact 11–12, 23, 124
  of reason 11–12, 23

Uncertainty 200
Universals 114
Unobservable 217

Wave picture 185–186

Zeno's paradoxes
  classical 82
  quantum 203–205

534112

DATE DUE

M[illegible] 1997

SEP 0 5 1997

CARR McLEAN, TORONTO FORM #38-297

BD 161 .B86 1983 no. 3
Bunge, Mario Augusto. 010101 000
Philosophy of science and tech
0 1163 0013170 7
TRENT UNIVERSITY

BD161 .B86 1983 no.3 pt.1
Bunge, Mario Augusto
Philosophy of science and technology
672491

DATE	ISSUED TO

672491

TREATISE ON BASIC PHILOSOPHY

Volume 8

ETHICS:
THE GOOD AND THE RIGHT

TREATISE ON BASIC PHILOSOPHY

1

SEMANTICS I *Sense and Reference*

2

SEMANTICS II *Interpretation and Truth*

3

ONTOLOGY I *The Furniture of the World*

4

ONTOLOGY II *A World of Systems*

5

EPISTEMOLOGY & METHODOLOGY I *Exploring the World*

6

EPISTEMOLOGY & METHODOLOGY II *Understanding the World*

7

EPISTEMOLOGY & METHODOLOGY III *Philosophy of Science & Technology*

Part I *Formal and Physical Sciences*
Part II *Life Science, Social Science and Technology*

8

ETHICS *The Good and the Right*

MARIO BUNGE

Treatise on Basic Philosophy

VOLUME 8

Ethics:

THE GOOD AND THE RIGHT

D. REIDEL PUBLISHING COMPANY

A MEMBER OF THE KLUWER ACADEMIC PUBLISHERS GROUP

DORDRECHT / BOSTON / LANCASTER

Library of Congress Cataloging-in-Publication Data CIP

Bunge, Mario Augusto.
Ethics: the good and the right / Mario Bunge.
p. cm. — (Treatise on basic philosophy: v. 8)
Bibliography: p.
Includes index.
1. Ethics. 2. Social ethics. I. Title. II. Series: Bunge, Mario Augusto. Treatise on basic philosophy; v. 8.
BJ1012.B75 1989
170—dc19 89-2394
ISBN-13: 978-90-277-2840-1 e-ISBN-13: 978-94-009-2601-1
DOI: 10.1007/978-94-009-2601-1

Published by Kluwer Academic Publishers,
P.O. Box 17, 3300 AA Dordrecht, The Netherlands.

Kluwer Academic Publishers incorporates the publishing programmes
of D. Reidel, Martinus Nijhoff, Dr W. Junk and MTP Press

Sold and distributed in the U.S.A. and Canada
by Kluwer Academic Publishers,
101 Philip Drive, Norwell, MA 02061, U.S.A.

In all other countries, sold and distributed
by Kluwer Academic Publishers Group,
P.O. Box 322, 3300 AH Dordrecht, The Netherlands.

All Rights Reserved

© 1989 by Kluwer Academic Publishers
Softcover reprint of the hardcover 1st edition 1989
No part of the material protected by this copyright notice
may be reproduced or utilized in any form or by any means,
electronic or mechanical, including photocopying, recording or
by any information storage and retrieval system, without
written permission from the copyright owner

GENERAL PREFACE TO THE *TREATISE*

This volume is part of a comprehensive *Treatise on Basic Philosophy.* The treatise encompasses what the author takes to be the nucleus of contemporary philosophy, namely semantics (theories of meaning and truth), epistemology (theories of knowledge), metaphysics (general theories of the world), and ethics (theories of value and right action).

Social philosophy, political philosophy, legal philosophy, the philosophy of education, aesthetics, the philosophy of religion and other branches of philosophy have been excluded from the above *quadrivium* either because they have been absorbed by the sciences of man or because they may be regarded as applications of both fundamental philosophy and logic. Nor has logic been included in the *Treatise* although it is as much a part of philosophy as it is of mathematics. The reason for this exclusion is that logic has become a subject so technical that only mathematicians can hope to make original contributions to it. We have just borrowed whatever logic we use.

The philosophy expounded in the *Treatise* is systematic and, to some extent, also exact and scientific. That is, the philosophical theories formulated in these volumes are (*a*) formulated in certain exact (mathematical) languages and (*b*) hoped to be consistent with contemporary science.

Now a word of apology for attempting to build a system of basic philosophy. As we are supposed to live in the age of analysis, it may well be wondered whether there is any room left, except in the cemeteries of ideas, for philosophical syntheses. The author's opinion is that analysis, though necessary, is insufficient — except of course for destruction. The ultimate goal of theoretical research, be it in philosophy, science, or mathematics, is the construction of systems, i.e. theories. Moreover these theories should be articulated into systems rather than being disjoint, let alone mutually at odds.

Once we have got a system we may proceed to taking it apart. First the tree, then the sawdust. And having attained the sawdust stage we should move on to the next, namely the building of further systems. And this for three reasons: because the world itself is systemic, because no idea can become fully clear unless it is embedded in some system or other, and because sawdust philosophy is rather boring.

The author dedicates this work to his philosophy teacher

Kanenas T. Pota

in gratitude for his advice: “Do your own thing. Your reward will be doing it, your punishment having done it”.

CONTENTS OF ETHICS

GENERAL PREFACE TO THE *TREATISE* v
PREFACE TO *ETHICS* xiii
ACKNOWLEDGEMENTS xv
INTRODUCTION 1
  1. Value, Morality and Action: Fact, Theory, and Metatheory 1
  2. Basic Schema of Values, Norms and Actions 3
  3. Relations between Axiology, Ethics and Action Theory 5
  4. The Task 6

PART I VALUES

1. ROOTS OF VALUES 11
  1. Nature and Kinds of Value 13
    1.1 Nature of Value 13
    1.2 Kinds of Value 16
    1.3 Summary 18

  2. Value Sources 19
    2.1 Biovalue 19
    2.2 Psychovalue 23
    2.3 Sociovalue 29
    2.4 Summary 33

  3. Needs, Wants, and Values 34
    3.1 Basic Needs and Legitimate Wants 34
    3.2 Values 36
    3.3 Value Categories and Orders 38
    3.4 Summary 40

2. WELFARE 41
  1. Welfare and Happiness 42
    1.1 From Bentham to Decision Theory 42
    1.2 Definitions 44
    1.3 Postulates 47

1.4 Social Welfare 49
1.5 Summary 52

2. Value Conflict and Change 53
2.1 Value Conflict 53
2.2 Ideals 55
2.3 The *Summum Bonum* 59
2.4 Summary 60

3. VALUE THEORY 61
1. Analysis 62
1.1 Intrinsic and Instrumental 62
1.2 Absolute and Relative 64
1.3 Objective and Subjective 67
1.4 Individual and Social 69
1.5 Fact and Value 71
1.6 Summary 73

2. Value and Knowledge 73
2.1 Semantic and Epistemological Status 73
2.2 Values and Science 75
2.3 Values and Technology 78
2.4 Summary 79

3. Value Measures and Calculi 80
3.1 Objective Value of a Generic Item 80
3.2 Objective Value of Systems and Processes 83
3.3 Subjective Value (Utility) 86
3.4 Value Calculi 88
3.5 Summary 89

PART II MORALS

4. ROOTS OF MORALS 93
1. Rights and Duties 95
1.1 Right and Duty 95
1.2 Rights Imply Duties 101
1.3 Summary 105

2. Morals 106
2.1 Right and Wrong Actions 106
2.2 Moral Problems 111

2.3 Morals 115
2.4 Summary 118

3. Sources and Functions of Morals 119
3.1 Biological 119
3.2 Psychological 122
3.3 Social 128
3.4 Summary 132

5. MORALITY CHANGES 133
1. Development and Evolution 134
1.1 Nature and Nurture 134
1.2 Origin and Breakdown of Norms 139
1.3 Summary 142

2. The Human Condition 143
2.1 Human Nature 143
2.2 Competition 146
2.3 Cooperation 148
2.4 Summary 151

3. Morality Changes 151
3.1 Evolution of Mores and Morals 151
3.2 The Current Moral Crisis 154
3.3 Summary 156

6. SOME MORAL ISSUES 158
1. Private Morals 159
1.1 Three Moral Spheres 159
1.2 Life and Death 162
1.3 Virtue and Sin 166
1.4 Ten Virtues 170
1.5 Desert and Reward, Crime and Punishment 176
1.6 Summary 178

2. Public Morals 179
2.1 Human Survival Issues 179
2.2 Equality 180
2.3 Freedom 184
2.4 Justice 187
2.5 Democracy 190
2.6 Summary 192

PART III ETHICS

7. TYPES OF ETHICAL THEORY 197

1. Matters of Method 199

1.1 Religious and Secular Doctrines 199
1.2 Monism and Pluralism 201
1.3 Absolutism and Relativism 204
1.4 Objectivism and Subjectivism 205
1.5 Emotivism, Intuitionism and Cognitivism 207
1.6 Consequentialism and Deontologism 209
1.7 Individualism, Holism, Systemism 212
1.8 Conservatives and Reformists 216
1.9 Summary 217

2. Moral *isms*: Egoistic 218

2.1 Nihilism 218
2.2 Rational Egoism 220
2.3 Libertarianism 223
2.4 Contractualism 226
2.5 Negative Utilitarianism 229
2.6 Summary 230

3. Moral *isms*: Altruistic 231

3.1 Natural Law 231
3.2 Kant 233
3.3 Utilitarianism 236
3.4 Agathonism 241
3.5 Summary 242

8. ETHICS *ET ALIA* 243

1. Ethics and Knowledge 244

1.1 Ethics and Logic 244
1.2 Ethics and Epistemology 248
1.3 Ethics and Ontology 251
1.4 Ethics and Science 255
1.5 Ethics and Technology 258
1.6 Summary 263

2. Ethics and Action 264

2.1 Praxis and Ethics: Generalities 264
2.2 Bioethics 266
2.3 Nomoethics 270
2.4 Business Ethics 276

2.5 Political Ethics 278
2.6 Summary 283

9. METAETHICS 285
1. Scientific and Philosophic Ethics 286
1.1 Ethics as Science and as Philosophy 286
1.2 Scientific and Axiological Basis 290
1.3 Constraints on Moral Norms 292
1.4 Norms and Normative Systems 294
1.5 Moral Reasoning 299
1.6 Summary 303

2. Ethics and Reality 304
2.1 The Is-Ought Gap 304
2.2 Moral Truth 306
2.3 Applying Ethics 309
2.4 Testing Ethics 311
2.5 Summary 314

PART IV ACTION THEORY

10. ACTION 319
1. Individual Action 320
1.1 Philosophical Underpinnings 320
1.2 Human Action 323
1.3 Goals, Means, and Plans 330
1.4 Work 332
1.5 Summary 334

2. Collective Action 335
2.1 Private and Public Interests 335
2.2 Social Policies and Plans 340
2.3 Cooperation and Competition 343
2.4 Competitive Cooperation 347
2.5 Management 349
2.6 Summary 352

11. SOCIAL PHILOSOPHY 354
1. Environmental 356
2. Biological 362
3. Economic 367
4. Political 374

5. Cultural 383
6. Summary 387

12. VALUES AND MORALS FOR A VIABLE FUTURE 390
1. The Thirteen Horsemen of the Apocalypse 390
2. Old and New Values and Morals 392
3. A Survival and Development Morality 393
4. A Viable Social Order 395
5. Conclusion 398

BIBLIOGRAPHY 400

INDEX OF NAMES 416

INDEX OF SUBJECTS 421

PREFACE TO *ETHICS*

This book is about values, morals, and human actions. It is also about axiology (the study of value systems), ethics (the study of moral codes), and action theory. It is concerned with both private and public values, morals, and actions. In particular, it seeks to uncover the roots and functions (biological and social) of valuation and morality. As well, it attempts to sketch a value system, a moral code and a general plan of action that may help us tackle the dreadful problems of our time.

We live in dangerous times. For the first time this may be the last time. A thesis of this book is that we have been marching blindly to the brink lured by wrong values and guided by wrong morals. But it is also a thesis of this book that such wrongs can be righted through gradual global social reforms, and that we are still in time to do so.

If the above theses are true, then value theory, ethics and action theory are nowadays a matter of life and death rather than just subjects of academic interest. In other words, it has become vitally important to know not only what values, morals and action patterns are, but also which are the values and morals we should live by, and which actions we should take.

The classical philosophers are of little help to find plausible and useful answers to these questions, for they never faced the possibility of the extinction of the human species as a result of nuclear war or environmental degradation. Nor could they make use of contemporary social science to work on axiological, ethical or praxeological problems. Our predicament is unique, and so is our chance to extricate ourselves from it.

Before the 1960s most value theorists and moral philosophers used to dwell in an ivory tower: they specialized in metaethics, were generally indifferent to real life problems, and seldom committed themselves to any substantive views on values and morals. (Bertrand Russell was an exception, but few professional philosophers took him seriously. He dealt only with large issues and anyone could understand him. Worse, he was a maverick.)

This situation has changed dramatically over the past quarter-century: value theory and moral philosophy have never been as alive as nowadays. Haunted by the spectres of nuclear war, environmental degradation, and social injustice, value theorists and moral philosophers have descended in droves from the ivory tower to the agora. On the whole this descent to social reality has been healthy: there is less hair-splitting and pointless analysis, less rhetoric and hypocrisy. But at the same time there is also more shameless defense of the supreme axiological blunder — the worship of possessions — and the supreme moral vice — selfishness.

The revival of value theory and ethics can be attested to by anyone who bothers to peruse the philosophical journals published in the course of the last few years. This revival is particularly welcome at a time when philosophy as a whole is at a low ebb — so much so that some philosophers have proclaimed its death while others have taken leave of reason. But the current flourishing of ethics may be an indicator of the general crisis of modern civilization, for people do not usually reflect on problems about values and morals until they face them, and nowadays most of us face them daily by the dozen.

This is the last volume of my *Treatise on Basic Philosophy*, on which I started to work two decades ago. It is consistent with the previous volumes, in particular with the naturalistic, dynamicist, emergentist and systemist ontology, as well as with the realistic and ratioempiricist semantics and epistemology formulated therein. However, the present book may be read independently of its companions.

Finally an autobiographical note. I began writing on value theory and ethics nearly three decades ago (Bunge 1960, 1961, 1962a). When I planned this volume I thought that writing it would be plain sailing. I was counting on my calculus of value (1973, 1975) and on decision theory, which I had applied to a political problem (1973). Fortunately, before I started to work on this book I realized the impossibility of a general value calculus, and I became disillusioned with decision theory (see Vol. 7, Ch. 5, Sect. 5.2). These disappointments forced me to take a fresh look at values, morals, and actions. This task proved to be more formidable than anticipated — and, by the same token, more rewarding too. I hope that my second thoughts on values, morals and action are an improvement on my earlier ones.

ACKNOWLEDGEMENTS

I am greatly indebted to Frank Forman (political philosophy), Ernesto Garzón Valdés (ethics and legal philosophy) and Pierre Moessinger (psychology) for having criticized a draft of this book. However, I have stuck to many ideas which they reject.

I have also benefited from exchanges on axiological or ethical problems with many of my students and with some scholars, in particular Joseph Agassi, Carlos E. Alchourrón, David Blitz, Eugenio Bulygin, Héctor-Neri Castañeda, José M. Ferrater-Mora, Andrés J. Kálnay, Mihailo Marković, Luis Puelles, Miguel A. Quintanilla, Hernán Rodríguez-Campoamor, Fernando Salmerón, Tom Settle, and Paul Weingartner. And I am grateful to Armand Buchs, Jean-Pierre Imhoff and Henri Ruegg for their hospitality at the Faculté des Sciences of the Université de Genève during the 1986—87 academic year.

My debt extends to a number of persons whom I never thanked enough while they were alive. Among them are the value theorists and moral philosophers Risieri Frondizi and Alfred Stern, both of whom I befriended in 1944 when I launched the philosophical journal *Minerva*, and whose courageous lives have been a model.

My greatest debt is of course to my parents. My mother Mariechen, who did her best to keep me on a short Lutheran leash, served a prison term at age 62 for conspiring against a military dictatorship. My father Augusto — physician, sociologist, writer, and congressman — combined an intense love of life and respect for nature with devotion to the public good and passion for social justice and liberty. Among other books he wrote one on moral and social philosophy: *El culto de la vida* (1915). He taught me through example that politics need not be dirty, and that it ought to be the arm of morality.

INTRODUCTION

The purpose of this Introduction is to sketch our approach to the study of value, morality and action, and to show the place we assign it in the system of human knowledge.

1. VALUE, MORALITY AND ACTION: FACT, THEORY, AND METATHEORY

We take it that all animals evaluate some things and some processes, and that some of them learn the social behavior patterns we call 'moral principles', and even act according to them at least some of the time. An animal incapable of evaluating anything would be very short-lived; and a social animal that did not observe the accepted social behavior patterns would be punished. These are facts about values, morals and behavior patterns: they are incorporated into the bodies of animals or the structure of social groups.

We distinguish then the facts of valuation, morality and action from the study of such facts. This study can be scientific, philosophic or both. A zoologist may investigate the way an animal evaluates environmental or internal stimuli; a social psychologist may examine the way children learn, or fail to learn, certain values and norms when placed in certain environments. And a philosopher may study such descriptive or explanatory studies, with a view to evaluating valuations, moral norms, or behavior patterns; he may analyze the very concepts of value, morals and action, as well as their cognates; or he may criticize or reconstruct value beliefs, moral norms and action plans.

In other words we distinguish the following levels:

Value systems	Moral codes	Action patterns
Axiology (value theory) (*a*) scientific (*b*) philosophic	*Ethics* (moral theory) (*a*) scientific (*b*) philosophic	*Praxiology* (action theory) (*a*) scientifico-technical (*b*) philosophic
Meta-axiology: philosophy of value concepts and theories	*Metaethics*: philosophy of moral concepts and theories	*Metapraxiology*: philosophy of action concepts and theories

The scientific study of human value systems, moral codes and action patterns was initiated in the 19th century by social scientists. One important contribution of these scientists was to suggest some of the biological and social determinants of valuation, morality and behavior patterns. Another was to exhibit a great variety of value systems, moral codes and action patterns across cultures. A third was to show the existence of certain cross-cultural axiological, moral and behavioral invariants underneath that variety — which, not surprisingly, proved the legitimacy of the much-battered concept of human nature.

On the other hand the philosophies of value, morality and action began in classical Antiquity. At least from the time of Socrates on philosophers have grappled with substantive problems about values, morals and actions. In particular they sought goodness criteria and righteousness norms. Some of them, particularly Plato, Aristotle, Aquinas, Hobbes, Locke, Spinoza, Hume, Bentham, Mill and Marx conceived of value theory and ethics as having an intimate relation with social and political philosophy, hence with collective action, in particular social reform.

However, the classical philosophers were notoriously imprecise. For example, Socrates was confused about the good, and Plato about justice. The analytic philosophers, from Moore on, attempted to clarify the basic axiological and ethical concepts, and their successors extended this work to action theory. True, they seldom went farther than to propose definitions, and they never used any advanced formal tools. Still, they succeeded in persuading many philosophers that one should be careful in using such terms as 'good' and 'right'.

Regrettably, in their effort to cleanse axiology, moral philosophy and action theory, most analytic philosophers gutted them. Indeed they shifted from substantive issues to questions of linguistic or conceptual analysis, losing historical perspective in the process. They advised us to analyze, not to evaluate or moralize, let alone act: they told us that we should deal with the language of values, morals and actions, not with valuation, morality or action. As a result most analytic value theorists, moral philosophers and action theorists became gutless, i.e. unwilling or incapable of grappling with any of the axiological, moral or practical problems of our time. Consequently their writings became ineffectual and boring.

The new wave in value theory, moral philosophy and action theory, which was generated in the 1960s by such political events as the civil rights movement, the fight against poverty and the environmentalist and peace movements, changed all that. Without abandoning analysis, philosophers started to tackle real life issues.

To be sure, there is still much to be done in these fields, for some of the hotly debated issues, such as taxes and abortion, affect only a minority; most treatments of value, morality or action problems are piecemeal rather than being embedded in comprehensive philosophical systems; and many of the analyses are phony, particularly when they involve game theory with payoff matrices the entries of which are invented.

In this book we shall approach in an analytic fashion some real and serious issues concerning values, morals, and action. We shall not only analyze certain key concepts but shall also propose definite substantive principles, and shall attempt to organize these into a theory.

Our principles will not be approved of by those who prefer to remain closeted in academe, and much less by those who — perversely in our view — identify rationality with the exclusive pursuit of private interests, thus giving ethics a bad name. (The selfish do not need morality except as a fig leaf.) But our principles may be appreciated, though not necessarily shared, by people concerned about the current crisis in values and morals, and sensitive to such crucial issues as war, the environment, and social justice.

2. BASIC SCHEMA OF VALUES, NORMS AND ACTIONS

The title of this book sounds Platonic. But its contents are not, for — in line with the Aristotelian and naturalist traditions — it denies the separate existence of good and evil, as well as of right and wrong. Indeed, we treat 'good' and 'bad', 'right' and 'wrong' as adjectives not nouns. These words denote properties of entities or processes, in particular actions, not entities in themselves.

Moreover we adopt the view that the good and the right are sometimes objective and at other times subjective, but at all times relative to people, or other animals, and their circumstances. In particular we hold that animals, in particular people, evaluate, whereas societies do not. However, there are social values, i.e. properties of social groups, that are favorable to the welfare of such groups. But social values result from activities of individuals, in particular from their economic, political and cultural activities.

We also take the commonsensical view that all normal animals strive to attain or retain a state of well-being — which, however, is not the same for all. Consequently normal animals value positively, i.e. they find good, anything they need for their well-being and, in the first place, for their survival. Moreover we postulate that needs and wants — visceral, mental or social — are the very roots of values, and that the function of norms is to protect such values, i.e. to facilitate their realization. We also postulate that we are driven by our values and constrained by our norms, not only by external factors.

However, not all values are on the same footing. There are primary, secondary, and even higher order values, according to the level of needs or wants they originate in. Correspondingly there are basic rights and duties, namely those associated with basic values, and higher order — i.e. less important — rights and duties, i.e. those that correspond to higher order values. As with rights and duties, so with norms and actions: the basic ones are rooted to basic needs, whereas all the others originate in wants, some legitimate and others not. Even such towering desiderata as equality, fraternity, liberty and justice are only ancillary to the right to life and the duty to protect life.

In short, we posit the following schema, to be spelled out through the book.

BASIC NEEDS: SURVIVAL Food, water, shelter, rest, etc.	LEGITIMATE WANTS: WELL-BEING Love, security, activity, advancement, etc.
PRIMARY VALUES Whatever meets basic needs.	SECONDARY VALUES Whatever meets wants.
PRIMARY GOALS Realizing primary values.	SECONDARY GOALS Realizing secondary values.
BASIC RIGHTS AND DUTIES To pursue or help pursue primary goals. E.g., the right to work and the duty to work well.	SECONDARY RIGHTS AND DUTIES To pursue or help pursue secondary goals. E.g., to participate in the responsible management of one's workplace.
PRIMARY NORMS Enjoin us to exercise basic rights or discharge basic duties. Supreme norm: "Enjoy life and help live".	SECONDARY NORMS Enjoin us to exercise secondary rights or discharge secondary duties. E.g., to promote mutual help, equality, and freedom.
PRIMARY ACTIONS Directed at attaining primary goals and, if deliberate, guided by primary norms.	SECONDARY ACTIONS Directed at attaining secondary goals and, if deliberate, guided by secondary norms.

In short, we postulate that values and morals do not hover above everyday life but, on the contrary, originate in our daily struggle to stay alive and attain or retain a state of well-being. Only people who have already met their basic needs can afford to dream of loftier goals. And it so happens that nowadays most human beings can hardly meet their basic needs, let alone realize their aspirations. Worse, the economic greed and political ambition of a few have put the survival of our species at risk. Therefore any value theory and any ethical doctrine that overlooks these facts is at best useless, and at worse a diversion from our most pressing responsibilities.

In light of the preceding we submit that any good moral code, and any correct ethical doctrine, must satisfy the following desiderata:

(i) *axiological soundness*, i.e. presupposing a value system conducive to human survival and development;

(ii) *realism*: dealing with real-life moral conflicts;

(iii) *viability*: possibility of living up to — i.e. designed for normal people not for either saints or psychopaths;

(iv) *agathonist*: seeking the good, hence including the maxim "Enjoy life and help live";

(v) *selftuist*: neither selfish nor altruistic, but combining private and public interests;

(vi) *logical rationality*, i.e. consistency;

(vii) *epistemological soundness*: resorting to normal cognition and making no claim to either infallible intuition or revelation;

(viii) *testability*: capable of being judged by the consequences of the actions ruled by it;

(ix) *corrigibility*: capable of being altered, without losing its essential properties, in the light of new circumstances and new knowledge;

(x) *universality*: capable of being adopted anywhere by anyone.

We shall do our best to comply with this tall order. But before proceeding to do so let us look into the place of axiology, ethics and action theory in our philosophical system.

3. RELATIONS BETWEEN AXIOLOGY, ETHICS AND ACTION THEORY

Axiology is centrally concerned with the good, ethics with the right, and action theory with actions that are both efficient and right. Now, a right action is one that promotes the good, whereas a wrong one promotes evil. Hence the good is in every regard — conceptually, biologically and technically — prior to the right. Not that the ability to discriminate good from bad suffices to sort out right from wrong. Whereas the former ability is presumably possessed by all animals, only animals possessing a modicum of moral sensibility, intelligence, imagination and social experience can tell right from wrong and, moreover, plan right and efficient actions.

The upshot of the preceding for the relations among the disciplines concerned with the good and the right is clear. Axiology precedes ethics as well as the other disciplines concerning praxis, namely action theory, social (in particular political and legal) philosophy, and sociotechnology. Shorter: Axiology precedes practical philosophy and sociotechnology.

Kantians and legal positivists disagree: They define the good in terms of the right. So do emotivists, though for a different reason: According to them the good is only a matter of emotion, not cognition, and the right is one of convention, not interests. We reject these alternative views because they overlook the mechanisms of formation of our ideas of right and wrong and because they block any attempts to bring science and technology to bear upon the right conduct of human affairs.

The above considerations about the relations between axiology, ethics and action theory are embedded in the broader picture of the place of these disciplines in our philosophical system, and their relations with the sciences and technologies directly relevant to practical philosophy: See Figure 0.1.

4. THE TASK

We have set ourselves three tasks. The first is constructive: We shall propose certain views on value, morality, and action. Our second task is analytic: We shall analyze the key concepts occurring in our principles (postulates and rules). Our third task is critical: We shall examine critically a number of rival views.

In performing all three tasks we shall make use of some of the results obtained in preceding volumes of their *Treatise*, though no knowledge of them will be assumed. We shall also attempt to remain close to real life and, particularly, to the great issues of our time. In so doing we will depart from the tradition according to which philosophers only write

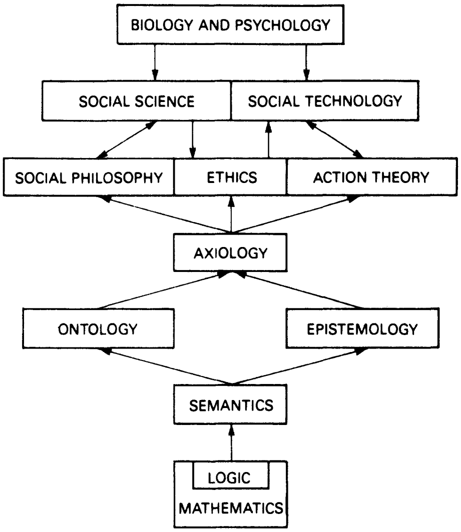

BIOLOGY AND PSYCHOLOGY
SOCIAL SCIENCE
SOCIAL TECHNOLOGY
SOCIAL PHILOSOPHY
ETHICS
ACTION THEORY
AXIOLOGY
ONTOLOGY
EPISTEMOLOGY
SEMANTICS
LOGIC
MATHEMATICS

Fig. 0.1. The place of axiology, ethics, action theory and social philosophy in the *Treatise*. Social philosophy encompasses economic, political, legal and cultural philosophy. The union of social philosophy with ethics and action theory may be called *practical philosophy* even though, like every other branch of philosophy, it is eminently conceptual. It is practical only in that it is concerned with deliberate human action, both descriptively and normatively.

for their colleagues. By the same token we shall be working in the tradition of the philosophers and social scientists concerned with improving the lot of their fellow human beings.

PART I

VALUES

CHAPTER 1

ROOTS OF VALUES

Values and morals do not exist by themselves, except for metaphysical idealists such as Plato. It is only for purposes of conceptual analysis that we may find it convenient to feign that they do. Actually there are only valuable or disvaluable objects — valuable, that is, for some organisms in certain states. When we abstract values from their bearers we make the same mistake as when we detach properties from the things possessing them. Likewise there are no morals in themselves. Instead, there are animals which, when behaving according to certain patterns, contribute to the welfare of other animals. We tend to abstract such patterns and call them 'moral norms'.

For instance, when saying that well-being is a biological and psychological value, we mean that, being a survival condition, we evaluate positively some states of physical health and psychological contentment: i.e. we say that well-being is good for us. Likewise when saying that honesty is a moral and social value we mean that we assign honest behavior, nay, honest people, a positive role in social life as well as in keeping our own peace of mind. No organisms, no needs, hence no values. No society, no social behavior, hence no social values, whence no need for moral norms.

In other words, values and morals have biological and social roots: See Figure 1.1. And, no matter how high we climb the tree of abstraction, those roots are still there — unless the tree happens to be dead, which is not uncommon for value theories and moral philosophies. We need to gain some knowledge of such roots in order to satisfy our curiosity concerning the modes of emergence and submergence of values and morals. But we also need such knowledge in order to justify (validate) value claims and moral norms — at least if we are rational enough to value such justification instead of surrendering to such dogmatic stands as nihilism, authoritarianism, emotivism, or intuitionism.

In this chapter we shall explore briefly the main roots of values, and in Ch. 4 those of morals. Our explorations will be brief for two reasons. Firstly, a detailed and serious study of the biopsychological and social

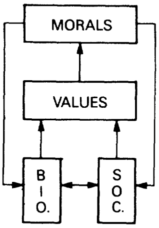


Fig. 1.1. The biological (visceral and mental) and social (economic, political and cultural) roots of values and morals. Example: We value food because it is necessary for survival, and companionship because it is part of our belonging to a society. The two roots, though distinct, are not mutually independent. For example, we often like to eat in company, which company may elicit some changes in the food we choose and the manner in which we eat it.

roots of values and morals behooves biologists, ethologists, psychologists, anthropologists, sociologists, economists, political scientists, and historians, rather than philosophers. Secondly, so far most of these scientists have paid little if any attention to values and morals, because they have been shackled by the philosophical superstitions — common to empiricism and intuitionism — that values and morals are subjective, and that subjective experience lies beyond the ken of science. In fact, they have been told by influential philosophers that there is an unbridgeable chasm between fact (the domain of science) and value (the domain of philosophy and art). Consequently they have believed that they should confine their attention to what is, without meddling into what ought to be. No doubt, this superstition was initially useful to separate the scientific grain from the ideological chaff in the study of society. But now it has become an obstacle to further advances in the understanding of man and in the design of policies and plans aiming at human betterment. Indeed, there is nothing science and technology (in particular sociotechnology) can do to mitigate human misery if they ignore what we ought to value and how we ought to behave.

In this chapter we shall study the nature, kinds, and sources of value. Some of the problems concerning the nature of value and value judgments will be studied at greater depth and length in subsequent chapters.

1. NATURE AND KINDS OF VALUE

1.1 *Nature of Value*

In an idealist metaphysics values exist by themselves above and beyond material things. A physicalist (or vulgar materialist) world-view makes no room for values except perhaps as subjective desires. In an emergent materialist perspective (Vol. 4 and Bunge 1981) there are no values in themselves but there are valuable items wherever there are organisms. According to this view values emerged on our planet about four billion years ago together with the first organisms capable of discriminating what was good for them, i.e. what was favorable to their survival, from what was not.

In the real world there are no values in themselves, anymore than there are shapes or motions in themselves. Instead, there are organisms that evaluate certain things (among them themselves) when they, as well as the things valued, are in certain states or undergo certain changes. In other words, whatever is valuable is so for some organisms in certain states, particularly states of deprivation that originate drives which motivate action.

Values are then relational or mutual properties, in the same boat with separation and adaptation, rather than intrinsic properties such as number of components and chemical composition (Bunge 1962a). The relational view of values contrasts with the absolutist view according to which values either exist by themselves or inhere in the things valued — a view espoused by Moore (1903) and others. Value absolutism is mistaken because it overlooks the facts that (*a*) there are no values without evaluating organisms, (*b*) values change with the internal and external circumstances of the evaluator, and (*c*) human value systems are culture-bound and therefore historically changeable — so much so that we approve of many items our parents disapproved of, and conversely. (However, some values are cross-cultural, i.e. they occur in all viable value systems. We shall come back to this point in Ch. 2, Sect. 2.1.)

Value theory is of course centrally concerned with value judgments. From a logical point of view a value judgment is a value of a valuation predicate for a given individual or *n*-tuple of individuals. Thus "Milk is good for Baby" results from applying the predicate "good" to the ordered pair ⟨milk, Baby⟩. But then all predicates, whether or not they are valuation predicates, can be analyzed as functions from collections of n-tuples of individuals to sets of propositions (Vol. 1, Ch. 1, Sect. 1.3).

"Good", a typical valuation predicate, can be applied not only to individuals, such as this book, but also to predicates, as when we say that altruism is good. In the latter case the evaluation predicate is of the second order: it is a property of a property, and therefore it comes under the sway of the higher order predicate calculus. For example, "good teacher" qualifies "teacher", a first order predicate, and the two together constitute a second order predicate. But there is also a first order analysis of the predicate in question. Indeed, by splitting the set of all teachers into the competent (C) and the incompetent ($\bar{C}$) ones, we succeed in analyzing "Liz is a good teacher" as "Liz belongs to C". This disposes of the intuitionist claim that "good teacher" (and for that matter every evaluation predicate) is unanalyzable. It also refutes the ordinary language philosophy thesis that there is a grammatical difference between "good" and "teacher": that whereas 'teacher' is an ordinary or predicative adjective, 'good' would be a sui generis or attributive one (Geach 1956, Williams 1972). There is no logical or grammatical difference between evaluation predicates and descriptive predicates.

A preliminary analysis of value judgments (Bunge 1962a) yields the result that the simplest among them are of the form "a is valuable for b", or "b values a". (The former exemplifies an objective value statement, the latter one that may be subjective.) A closer examination is bound to disclose complexity: it is likely to result in formulas of the type: a is valuable in respect b for organism c in circumstance d, where the word 'respect' is a synonym for 'property' or 'feature', and the term 'circumstance' denotes the states of both the object a, or value-bearer, and the subject c of valuation. A finer analysis may yield further relata (variables), as e.g. in

> a is valuable in respect b for organism c in circumstance d with goal e and in the light of the body of knowledge f.

In short, value judgments involve at least binary relations: they are of the forms Vab, $Vabc, \ldots, Vabc \ldots n$. In the case when we succeed in quantitating values, the relation becomes a function from n-tuples of objects to numbers. Example: $V(a, b, c, d, u) = v$, where u is a suitable unit, and v the numerical value c attributes a in respect b and circumstance d. In other words, the general form of a numerical value function is

$$V: A \times B \times \ldots N \times U \to \mathbb{R},$$

where A is a collection of value bearers, i.e. of valuable or disvaluable objects, B a collection of organisms, and the remaining factors in the cartesian product, up to N, may be collections of things, properties, states, or processes (in particular brain processes), whereas U is a set of units, and $\mathbb{R}$ the set of real numbers. However, quantifiable values are exceptional.

Note the following points. Firstly, values are not things, states of things, or processes in things: these can only be value-bearers of objects of valuation. Values are *relational* (or *mutual*) *properties* attributed to objects of certain kinds by organisms of certain types and in certain states. However, we may abstract values from their bearers and speak of *truth* (instead of true propositions), *beauty* (instead of beautiful things or processes), *fairness* (instead of fair actions or persons), and so on. We may do it in order to gain in generality and as long as abstraction is not mistaken for reification. Such a mistake, namely the idea that values (and the corresponding norms) exist by themselves, may be dubbed the *idealistic fallacy.*

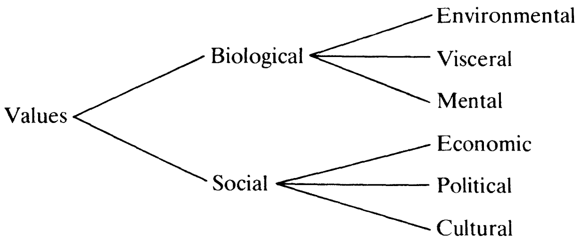
Secondly, because values are relational properties, there are no intrinsic or absolute values representable by unary predicates such as "alive" or "bankrupt". A phrase such as 'a is valuable', or 'Va' for short, is incomplete: it must be interpreted as an abbreviation for 'a is valuable in some respect x for some organism y in some circumstance z'. Thirdly, because "relational" is not the same as "subjective" (Vol. 7, Ch. 2, Sect. 3.1), the thesis that values are relational properties differs from the view that all values are subjective, such as the feelings of pride and shame. Only some values, particularly the aesthetic ones, are subjective, in the sense that there are hardly any objective standards enabling us to judge the corresponding evaluations.

Whether objective (like nutritive value) or subjective (like happiness), all values are *secondary* properties for depending on some animal subject or other. In this regard they are similar to color and loudness, and quite different from primary properties such as wavelength, sound intensity, and chemical or social composition. (Recall Vol. 3, Ch. 2, Sect. 2.2.) Values may also be regarded as systemic or *emergent* properties in that they are properties of systems that their components (the valued object and the evaluating subject) lack when in separation. (Frondizi 1963 called them *gestalt* properties.) In sum, our view of values is relativistic but not subjectivistic. More in Ch. 3, Sect. 1.3.

Finally, note that we have placed no restriction on the kind of objects that can be evaluated, i.e. on the value-bearers. In principle anything can be attributed a value: thing, state, event, process, feeling, disposition, experience, concept, procedure, or what have you. Not only "positive" items such as things and states of things, but also "negative" ones such as their absence or disappearance can be the object of valuation. For example, the loss of life and the absence of educational facilities are generally regarded as disvaluable. They are so regarded because the corresponding "positive" objects, e.g. life and educational facilities, are deemed to be good. Likewise the end of a war, or of political persecution, are regarded as good because the processes themselves are attributed a negative value.

1.2 *Kinds of Value*

We shall partition the collection of all values into two large families: biological and social. In turn, we divide biological values into environmental (e.g. clean air), visceral (e.g. adequate food), and mental (e.g. feeling needed). And we admit three genera of social value: economic (e.g. productivity), political (e.g. self-government), and cultural (e.g. the advancement of knowledge). In turn, every genus may be split into several species. For instance, the cultural values may be grouped into cognitive, moral, and aesthetic. In short, we adopt the following value divisions:



Some objects are polyvalent, i.e. they are valuable or disvaluable in more than one respect. For example, a loaf of bread is both biologically and economically valuable. A piece of jewelry is psychologically, economically and perhaps also aesthetically valuable but, since it meets no biological need, its visceral value is nil. A political rally has a political value and, indirectly, a psychological and perhaps an economic one as well. Because one and the same object may be valuable (or disvaluable) in more than one way, it is not possible to partition the collection of all value-bearers into mutually disjoint classes. Hence, although we can class *values* (or value functions), we cannot class the value-*bearers* other than in biologically or sociologically meaningful ways.

The above tripartite division of social values into economic, political and cultural parallels our analysis of society into its economy, polity, and culture (Vol. 4, Ch. 5, Sect. 2.4). Note that admitting that there are different kinds of social value does not entail their mutual independence. In particular the moral values, which may be included among the cultural ones, are not independent of the economic values, which in turn are (or ought to be) constrained by certain moral norms. For example — as a character of G. B. Shaw said — the destitute cannot afford to be honest; and productivity ought not to be sought at the cost of welfare.

Again, the tripartite division of cultural values into cognitive, moral and aesthetic does not entail their mutual independence. For example, it takes some learning to appreciate aesthetic values, and some moral courage to expose fake science. Likewise it usually takes some knowledge, in addition to some moral sensibility, to realize that certain problems are of the moral kind, and far more knowledge and moral sensibility are needed to come up with correct solutions to them. Still, the mutual dependence of values of different types does not prevent them from being distinct. (This exemplifies a logically necessary platitude: Y depends upon X if and only if (a) X is not the same as Y, and (b) some changes in X are accompanied by changes in Y.)

The point of distinguishing and classing values is not just to bring some tidyness into the fuzzy domain of value theory. The main point is to avoid attempting to compare values that are non-comparable, such as those of a miner's day work with writing a poem, proving a theorem, or playing a soccer match. A miner's day work may be compared with a clerk's or a manager's, a poem with another poem, and so on. This is obvious, yet it contradicts a basic tacit assumption of utilitarian philosophy, namely that all items, irrespective of their type, can be compared and ranked.

Even if we succeeded in constructing one numerical utility function for every evaluator, it would not do to apply it to objects belonging to disjoint categories. And yet the assumption that a single utility (or subjective value) function suffices for everything with respect to any person, is a basic postulate of mainstream economics, decision theory, and contemporary utilitarianism. Granted, the postulate is not too far-fetched in mercantilist societies, where nearly everything can be bought and sold — even loyalty and affection. But the point is that moral philosophers ought to indicate that this state of affairs is abnormal, hence it ought to be corrected. In a rational and just society not everything can be traded for something else: some things are inalienable, and only some things can be traded for certain other things. Shorter: Price, though important in the market place, is irrelevant to values and morals. We shall return to this question in Sect. 3.1.

So much, for the time being, for the variety of kinds of value. In what follows we shall attempt to justify the above partitions as well as to ferret out the peculiarities of each kind of value.

1.3 *Summary*

Values are not entities but properties of certain items and, more particularly, they are mutual or relational rather than intrinsic properties. More precisely, values are relations between certain objects (things or processes) and organisms. The former are the value bearers, the latter their (actual or potential) users. In the case of humans, values fall into two main categories: biological and social. In other words, whatever we value, we do so either because it meets a biological or a social need or want.

Things of one and the same kind can be evaluated in a number of different respects. For example, a mountain may have both an economic and an aesthetic value, and neither is reducible to the other; again, an album of junk music may fetch a high price. In other words, we must distinguish various value relations and functions. This holds, a fortiori, for items of different kinds. For example, even if a bouquet of flowers and a dinner bear the same price tag, their aesthetic and biological worth are not the same: they are hardly comparable. In short, there is no such thing as a single value ranking, let alone a single value function.

2. VALUE SOURCES

2.1 *Biovalue*

So far we have taken it for granted that certain items are valuable to organisms of certain kinds when in certain states. In this and the following sections we shall ask *why* certain objects are valued. It won't do to reply that we value whatever we desire, because we must reckon not only with wants but also, nay mainly, with basic needs, whether or not we happen to wish to meet them. (In fact we shall postulate that what makes something valuable is its ability to meet some need or want, regardless of whether the object in question is scarce. For similar approaches see Hull 1944, Handy 1962, and Parra-Luna 1983.)

Nor will it do to ask the economist for an answer to our question, for several reasons. Firstly, economists restrict the category of valuable items to commodities, i.e. goods and services; they are not interested in non-marketable items, such as clear skies and neighborly feelings, except perhaps as possible stimuli or inhibitors of production or consumption. Secondly, they limit their attention to the exchange value or price of commodities regardless of their objective biological or social value in use. Third, even so they still owe us a true theory of the evaluation or pricing mechanism. All that mainstream economists tell us is that commodity X is worth P dollars if and only if the buyers in an ideally free (purely competitive) market are prepared to pay P dollars for X. (Hence, if X finds no buyers, X is declared worthless.) But this answer dodges the question whether P is fair, excessive, or too low. In addition, it ignores the price-tagging mechanism, which is not exclusively a function of demand, and sometimes involves dumping, collusion, and other immoral practices. In short, economics is not in a position to tell us why we value certain objects. Far from investigating this problem, it postulates that everybody values something. Worse, most neoclassical economists postulate, without a shred of empirical evidence, that every person can rank the items in an arbitrary collection, or even assign them numerical utilities. (Recall Sect. 1.2.)

Let us move from social to natural science. Biology employs more or less explicitly a concept of value that may be regarded as the most basic, though not the only, axiological concept. This concept may be defined as follows:

DEFINITION 1.1 If a is a feature (organ, property, process, etc.) of an organism *b*, or of the environment *e* of *b*, then *a is valuable to b in e* $=_{df}$ the possession of, or access to, *a* favors the survival of *b* in *e*. Otherwise *a* is either neutral or disvaluable to *b* in *e*.

The valuable or disvaluable feature in question may help or hinder the organism in either of two ways: internally (physiologically) or externally (environmentally). In the former case what improves or worsens is the health of the organism in the given environment. In the latter case the feature in question improves or worsens the chances of survival of the organism when involved in interactions with either conspecifics or members of other species. Accordingly we distinguish *physiological* from *selective* value. At first sight these two are not independent: it looks as though physiological value were to ensure environmental or selective value. Not so, for a perfectly healthy organism may fall prey to an organism of another species.

Definition 1.1 refers to individuals: it identifies biological value with that which favors the survival of the individual. It makes no reference to the biospecies to which the individual belongs, or to the population of which it is a part. Hence it makes room for features that are unique to the individual in question, and therefore give it an edge or a handicap. To be sure, biologists often talk about something being valuable or disvaluable to a biospecies. This is a mistake, for biospecies are not concrete things but collections of such, and collections are concepts, hence they are not in the survival game. (Recall Vol. 4, Ch. 3, Sect. 1.3, or Vol. 7, Part 2, Ch. 3, Sect. 2.2.) Therefore the phrase '*X* is valuable to biospecies *Y*' must be understood as short for '*X* is valuable to every member of the biospecies *Y*', or as '*X* favors the survival and reproduction of members of biospecies *Y*'.

Our concept of biovalue (Definition 1.1) differs from that of biological (or Darwinian) fitness, which occurs in evolutionary biology. In the latter "fitness" is defined as survival or adaptive value, which in turn is equated with fecundity or, more precisely, with the size of the offspring that an organism of a species reproduces faithfully ("true to" the defining features of the species). We find no use for this notion in value theory for several reasons. Firstly, it has yet to be exactified in a manner that commands the consensus of the scientific community. Secondly, fertility is an indicator of fitness not a synonym for it. Thirdly, the equation of fitness with fertility presupposes an environment with unlimited resources, where a biopopulation can increase indefinitely in numbers provided it prevails over its competitors. But as a matter of fact every natural environment is limited; hence, by reproducing fast and in great numbers, the organisms of a given species may soon deplete and pollute their niche and thus bring their own extinction upon themselves. A high fertility or Darwinian fitness may therefore be fatal to a biospecies. The optimal fertility value of a biopopulation is, except in the cases of endangered species, well below the maximum. This is particularly clear in the case of humankind, which is suffering acutely from overpopulation. Such excessive reproductive success is threatening the very viability of humankind: though valuable up until a couple of centuries ago, it has now become disvaluable for having resulted in a world population that is far above the carrying capacity of our planet. However, let us go back to our theme.

Whatever favors the survival of an organism, hence is biovaluable to it, contributes to its health; conversely, whatever is detrimental to its health is biologically disvaluable to it even though it may be valuable in other respects. We must then supplement Definition 1.1 with a definition containing explicitly the concept of health of an organism of a given species. This concept was introduced by Definition 3.6 in Vol. 4. The gist of it was that health is the complement of sickness, which in turn equals the relative departure of an organism's functions from their physiological optimal values. All this was relativized to the environment, for different environments are likely to shift the values of both the actual and the normal (or optimal) functions.

The interest of all this for value theory is that it supplies an objective quantitative measure of biological value. The intuitive idea is this. Let a be a feature of an organism b or of its environment e. Then the value of a to b in e equals the degree of health of b in e when equipped with a or (in case a is a feature of e) when having access to a, minus the degree of health of b in e when deprived of a. More precisely, we propose the following slight modification of Definition 3.9 in Vol. 4:

DEFINITION 1.2 Let $H_A : B \times E \times T \rightarrow \mathbb{R}$ be the health function for organisms of kind B, in environments of type E, when the B's are in possession of, or have access to, features of kind A. Similarly, call $H_{\bar{A}} : B \times E \times T \rightarrow \mathbb{R}$ the corresponding health function of B's in E's when the A's are missing. In both cases, T is a subset of the real line $\mathbb{R}$, and every element t of T is interpreted as an instant of time. Then *the value of A for B in E* is the function $V: A \times B \times E \times T \rightarrow \mathbb{R}$ such that, for any a in A, any b in B, and any t in T,

$$V(a, b, e, t) = H_A(b, e, t) - H_{\bar{A}}(b,e,t),$$

which is *the value of feature a for organism b in environment e at time t.*

Since health ranges between 0 and 1, biological values lie between -1 and 1. In particular,

a is *maximally valuable* to b in e at t iff $H_A(b, e, t) = 1$ and $H_{\bar{A}}(b, e, t) = 0$;

a is *worthless* to b in e at t iff $H_A(b, e, t) = H_{\bar{A}}(b, e, t)$;

a is *maximally disvaluable* to b in e at t iff $H_A(b, e, t) = 0$ and $H_{\bar{A}}(b, e, t) = 1$.

So much for our first definitions. Next we make two assumptions so obvious that they are hardly ever stated explicitly:

POSTULATE 1.1 All organisms possess features (components, states or processes) that are valuable, and others that are disvaluable, in some environments.

POSTULATE 1.2 All environments have features that are valuable, and others disavaluable, for organisms of some species.

Note that the first axiom, though a platitude, is not a tautology. Moreover, from a radical reductionist viewpoint — which is not ours — it makes no sense to attribute values to components, states or changes of state of an organism, let alone to its environment, particularly to the physical components of the latter. Nor is the second axiom tautologous, for it draws our attention to the exceptional fitness of our terrestrial environment — hence to the need to protect it from our own depredations.

We are now ready for the key convention in this section:

DEFINITION 1.3 Let x be an item (thing or process) in organism b or in the environment of b. Then x is *healthwise good* for b if and only if x contributes to keeping b in good health.

An equivalent definition is this: Whatever contributes to the homeostasis (approximate constancy or balance of the *milieu intérieur*) of an organism is healthwise good for the latter. Shorter: Homeostasis "serves" health, hence self-preservation, whence it is biologically good for the organism in question. (No teleology here: Those organisms whose homeostatic mechanisms fail, whether for internal or external causes, get sick or even die.)

Note that ours is not a definition of "good" but of the narrower concept of good for health (or survival). Only very sick people, like Nietzsche, would claim that health alone counts and would consequently propose a biologistic value theory — one in which even truth would be reduced to survival value. Normal people know that there are nonbiological (e.g. cultural) values, and that sometimes biological needs must be postponed for others. Certainly, *primum vivere, deinde philosophari*, but human life is not only eating and drinking: it is also caring for others, learning, playing, and much more.

Finally, a few words about the nature/nurture controversy in relation to values. We are born liking and desiring certain things, such as warm milk and body contact, and disliking and rejecting others, such as feeling too warm or hearing loud noises. Such basic values arise from the corresponding basic needs, meeting which is necessary to stay healthy. But other values, in particular most of the social ones, are learned as we grow up, the same way we learn to walk or speak. And we may not learn such values as truth and beauty unless we happen to develop in an environment where truth and beauty are appreciated.

The thesis that all values are independent of experience, hence cannot be learned, is just as fantastic as the thesis that we are born with a substantial body of intellectual and liguistic knowledge, which experience can only channel or hone. Only a few basic values, first of all survival, "are in our genes". The healthy newly born craves mother's milk not steak or passion fruit; but given the opportunity he may acquire a taste for either. He is naturally active but does not go straight from the nursery to the soccer field or the ski slope. He is naturally inquisitive but at birth does not know anything, not even that knowledge is valuable. He is a natural evaluator, but he has to learn, often the hard way, the social values that will make it possible for him to belong to a certain social group.

To sum up this section: The source of biovalue is biological need. I.e., whatever an organism needs to keep healthy is good for it, at least on the biological side. This thesis will be generalized in Sect. 2.4 to read: The source of every value is some need or want. This is an example of what dons out of touch with life call the "naturalistic fallacy", following Moore (1903).

2.2 *Psychovalue*

So far we have dealt with organisms in general, claiming that certain biofunctions (such as the consumption or the synthesis of glucose), as well as certain environmental items (such as water and light), are objectively valuable for them. From now on we shall restrict our considerations to animals, for these alone are capable of evaluating some of their internal states, as well as some environmental states, as favorable, unfavorable, or indifferent to them, as well as to act accordingly. To put it anthropomorphically: Only animals can know (some of) what is good or bad for them. (But of course only to the higher animals may we attribute cognitive capacities: recall Vol. 4, Ch. 4, Sect. 4, or see Bunge 1980, or Bunge & Ardila 1987.) We posit then

POSTULATE 1.3 All animals have the ability to evaluate some of their own internal states and some environmental inputs, as well as the ability to act so as either to preserve or alter their own internal states, their environment, or both.

(Actually this hypothesis is too weak, for some bacteria — which are far below the animal kingdom, discriminate between beneficial and nocent stimuli of certain kinds, e.g. by moving up a nutrient gradient and down an acid one (chemotaxis); they may do this with the help of a molecular memory mechanism: Koshland 1977.)

Evaluation and motility allow animals to take action to improve their welfare or even to save their lives. They endow animals with a certain degree of freedom from some environmental pressures, as well as with some freedom to act in their own interests and, in the case of some higher animals, in the interests of their conspecifics as well. On the other hand plants and fungi, lacking valuation sensors as well as overall motility, are far more at the mercy of their environment. In short, the valuation ability has a great survival value.

(Physicalists may object to talking of values with reference to primitive organisms. They may argue that, after all, one might also claim that candles need, hence value, oxygen, or that electric bulbs need, hence value, electric current. But this would be biocentric talk, for candles and light bulbs last the less, the more they draw from their environment, whereas organisms live the longer, the more efficient their metabolism and, in the case of animals, also the better adapted their behavior. In any event candles and bulbs burn, do not live.)

The statement that an animal can evaluate items of certain kinds amounts to the proposition that it has a value system or set of preferences — e.g. that it prefers warmth to cold, or bananas to apples. This concept was elucidated by Definition 4.27 in Vol. 4; we reproduce it here in a somewhat improved form:

DEFINITION 1.4 Let S be a collection of items and b an animal capable of comparing them pairwise. Call S^* the partition of S into the family of equivalence classes, every one of which consists of equivalent objects (e.g. large ripe bananas). The ordered pair $\mathscr{V}_b = \langle S^*, \succsim \rangle$ is a *value system* of b in a given state in a given environment if, and only if,

(i) b can detect any member of S, discriminate it from all other items in S, and lump it together with its equivalents;

(ii) for any two members x and y of S^*, b either prefers x to y ($x \succ_b y$) or conversely ($y \succ_b x$) or neither ($x \sim_b y$) at a given time.

Belonging to an animal species involves, among other features, sharing some values with all conspecifics. That is, even though the individuals in a given biospecies may differ considerably from one another, and even though valuation is an individual affair, every animal species has a characteristic value system, no matter how idiosyncratic the value systems of its members may be. If they did not share some values they would not have the same basic needs, hence they would not belong to the same species. We elucidate the notion of a species-specific value system in terms of the individual preference relation $\succsim_b$ as follows.

DEFINITION 1.5 Let S be a collection of items and K an animal species the members of which are capable of comparing pairwise those items and of lumping together equivalent items. Call S^* the family of equivalence classes of S's. The ordered pair $\mathscr{V}_K = \langle S^*, \succsim_K \rangle$ is the K *value system* in a given environment and at a given time if, and only if, for any x and y in S^*,

$$x \succsim_K y \text{ if and only if } x \succsim_b y \text{ for all } b \text{ in } K$$

in the given environment and at the given time.

The main value of the last two definitions is that they set the stage for

POSTULATE 1.4 Every animal b has its own value system or preference ranking $\mathscr{V}_b$, and every animal species K has its own species value system $\mathscr{V}_K$.

This innocent-looking hypothesis has the consequence that there are universal human values, i.e. values shared by all humans regardless of their differences:

COROLLARY 1.1 The human species has a value system of its own.

Our next assumption is that animals only value environmental items insofar as these are instrumental in preserving or attaining certain internal states, in particular states of health. Thus ultimately *what is valued is being in certain states.* In other words, the preferences over environmental items are due to preferences over internal states (Bunge & Ardila 1987 Sect. 8.2.) This hypothesis can be formulated thus:

POSTULATE 1.5 Let S_K be the collection of internal states in which members of the animal species K can be, and S_E the collection of environmental items accessible to the K's. There is an injection φ: $S_E \to S_K$ from environmental items to internal organismic states such that, for every individual b of species K,

> if e and f are in S_E, then: $e \gtrsim_b f$ if and only if $\varphi(e) \gtrsim_b \varphi(f)$.

Why should an organism prefer to be in one state rather than in another? We take a clue from the neobehaviorist psychologist Hull (1944): b prefers state $s = \varphi(e)$ to state $t = \varphi(f)$ just in case the $t \mapsto s$ transition is accompanied by a reduction in a primary drive of b's. We adopt this hypothesis and turn it into

POSTULATE 1.6 If s and t are internal states of an organism b, then $s >_b t$ if and only if the transition from t to s is accompanied by a reduction in some need or want of b's.

By associating environmental items (things or processes) with their own internal states, and transitions between such states with need or want reductions, the animals endowed with learning abilities can learn new environmental item-internal state associations, and thus change their value systems. This occurs both in the course of individual development and of that of the evolution of animal populations. We compress this idea into

POSTULATE 1.7 All animal species alter their value systems in the course of evolution, and all individual animals can modify their individual preferences (or value systems) in the course of their lifetimes.

Now, learning is not always advantageous: in fact one often learns maladaptive behavior patterns, such as taking drugs or resorting to violence in the face of every conflict. For example, a rat will prefer receiving mild electric stimulation to its subcortical pleasure center, to eating (Olds and Milner 1954); and Rhesus monkeys given the choice between food and cocaine prefer the latter (Aigner & Balster 1978). These facts force us to distinguish *psychovalues* (psychological values) from *biovalues.* The biovalue of item a for animal b in a given state is the objective value that a possesses for b when in that state, whereas the psychovalue of a for b is the value that b *attributes* a — not always for its own good.

A discrepancy between a psychovalue and a biovalue may become so large as to constitute an internal conflict, even a tragedy. The existence of such internal conflicts was traditionally understood as the perpetual struggle between spirit (or soul or mind) and matter — except that in the cases of drug addiction, masochism, sadism, power addiction, and other vices, the allegedly immaterial spirit is manifestly far weaker than the material body. The scientific explanation proposed by biopsychology is quite different: Every conflict between a psychovalue and a biovalue is an instance of the dominance of one part of the brain (the one that learned the maladaptive behavior pattern) over the rest of the body.

The mere existence of discrepancies between values of the two kinds should warn us against the (operationist and behaviorist) temptation of inferring (objective) values from (manifest or "revealed") preferences — a common mistake among moral philosophers and economists. The two value systems must be distinguished even though both are ultimately rooted in biological processes. They are different because they occur in different parts of the body; in particular, psychovaluation is a physiological process occurring only in certain parts of the brain.

In any event, humans and other higher animals value certain items even if these do not meet some of their basic biological needs. Such animals have wants, desires, or even urges (compulsive desires) that may or may not match their biological needs, and thus may or may not contribute to their health. (For example, b may want x, which b needs to survive, as well as y, which is bad for b's health.) We take it that the concept of want, wish, or desire, can be explained scientifically in terms of drive, motivation, or pleasure, and use it to frame

DEFINITION 1.6 Let x be an item (thing or process) internal or external to organism b endowed with mental abilities. We stipulate that x is *psychologically valuable* for b in circumstance c if and only if b desires or wants x when c is the case.

Now, as noted above, not everything psychovaluable is good for us in all respects. Therefore we must distinguish between psychologically legitimate and illegitimate wants:

DEFINITION 1.7 If item x is psychologically valuable to animal b in circumstance c, then x is a *psychologically legitimate want* (or desire or wish or aspiration) of b in c if and only if x contributes to the long-term health of b. Otherwise x is *psychologically illegitimate.*

This is not a definition of a fully legitimate want, but only of a psychologically legitimate one: a full justification of a desire, as different from a mere whim, requires that it meets certain social requirements. However, this is a matter for the next section.

We can now tackle one of the oldest and most persistent problems in value theory, namely Plotinus's: Do we desire what is good, or do we call 'good' that which we happen to desire? Spinoza (1677, Part III, Comment on Prop. IX) held that we judge good whatever we desire, not the other way round. This is in fact the case with very young children: they judge all desirable things to be good — until they learn that some of them, though still desirable, are bad for their health or their social standing. Growing up and, in particular, the maturation of moral conscience, involves learning what is objectively good for ourselves and others, and putting up with burdens and pains for the sake of goodness. The upshot is that sometimes we desire what we believe to be good, and at other times we judge good what we desire. A perfect balance between these two inclinations is only attained when the good and the desired coincide. When they do not we face a value conflict. We postpone the study of value conflict to Ch. 2, Sect. 2.

So far we have not needed the notion of goal-seeking behavior: most animal behavior is automatic, even when it involves correct (i.e. life preserving) valuation. However, the higher animals, or at least the primates, are capable of performing voluntary (intentional) actions on the strength of prior evaluations. As conceived by physiological psychologists, the will is not an immaterial entity or a property of an immaterial mind, but a neural activity. Animal b wills action a if and only if b forms the purpose of doing a. By monitoring the firing of neurons preceding the performance of voluntary actions it has been possible to locate the will in the frontal lobes. (See e.g. Evarts, Shinoda & Wise 1984.) It makes then good scientific sense to acknowledge free will and characterize it by means of

DEFINITION 1.8 An animal acts of its own *free will* (or simply *freely*) if and only if its action is voluntary (or spontaneous) rather than either genetically programmed or compelled by external stimuli.

An animal endowed with a will of its own may, within bounds, choose to do what it wants. (That a particular choice may prove to be inconvenient to the individual or to others is another matter. If it does not have a fatal outcome, the animal may learn not to repeat the bad choice. For example, a single trial may teach a rat to avoid food of a certain kind.) The concept of choice, central to moral philosophy and social science, was characterized by Definition 4.28 in Vol. 4:

DEFINITION 1.9 Let $\mathscr{V}_b = \langle S, \gtrsim_b \rangle$ be a value system for an animal b at a given time (i.e. in a certain state), and call $A \subset S$ the set of alternative courses of action open to b at the time. We say that b *chooses* (or *selects*) option x in A if and only if

(i) it is possible for b to pick (i.e. do) freely any alternative in A (i.e. if b has both the ability and the opportunity of taking freely any of the courses of action in A);

(ii) b prefers x to any other options in A; and

(iii) b actually picks (i.e. does) x.

Choice patterns vary in the course of time due to changing circumstances (opportunities, needs, and wants) as well as of a result of learning. In particular the animal may learn, often by blind trial and costly error, to improve its choices, e.g. to seek or avoid the company of other individuals. But in all cases there is choice proper only if there is some freedom to act (or active freedom): without the latter the animal acts instinctively or on cue.

All of the higher animals enjoy *some* freedom, whence they are capable of *some* choices. But these are always bounded by ability and circumstance, particularly by environmental (natural or social) constraints. In some cases our choices are restricted to such dramatic options as starve or steal, flee or fight, shoot or be shot. The claim that we are or can become *totally* free to choose whatever we please is sheer theological or economic dogma. (See Vol. 7, Part 2, Ch. 4, Sect. 5.2.)

2.3 *Sociovalue*

In this section we shall refer exclusively to gregarious animals. For such animals sociality itself is valuable and enjoyable, for they could hardly survive on their own. In fact they need one another even when adult, and such need is the ultimate source of social values (such as peace) and, in the case of some animals, that of moral norms as well. The need for sociality is manifested in a variety of prosocial behavior patterns, such as parental care, tolerance, mutual help, and team work. And, among the social primates, prosocial behavior takes additional forms, such as division of labor and the erection, maintenance, reform, or dismantling of social organizations. On the other hand antisocial behavior, in particular selfishness, cheating, betraying, hoarding, and systematic violence against members of the same group, is disvaluable and consequently discouraged. Any society that tolerates frequent antisocial behavior is in for extinction. See Table 1.1.

Two of the concepts occurring in the preceding are clarified by

DEFINITION 1.10 The social behavior an individual engages in is

(i) *prosocial* if and only if it helps other members of the same society keep or improve their health, or meet their psychologically legitimate wants;

(ii) *antisocial* if and only if it hinders other members of its own society keep or improve their health, or meet their psychologically legitimate wants;

(iii) *socially neutral* if and only if it is neither prosocial nor antisocial.

TABLE 1.1. Examples of prosocial and antisocial feelings, attitudes and actions. Note the distinction between moral and nonmoral items. And note that some of them are not mutually exclusive. Thus there are love-hate and cooperation-competition relationships, mock fighting, and lying for the protection of the weak.

Prosocial	Antisocial
Moral	
Benevolence	Malevolence
Compassion	Cruelty
Concern for others	Indifference to others
Cooperation	Defection
Empathy	Anempathy
Fairness	Unfairness
Helping	Attacking
Loyalty	Betrayal
Reciprocity	Selfishness
Sharing	Hoarding
Nonmoral	
Admiration	Contempt
Friendship	Enmity
Gregariousness	Withdrawal
Hospitality	Inhospitality
Love	Hate
Parental care	Parental neglect
Play	Fight
Promise	Threat
Sympathy	Antipathy
Team work	Solitary work

Prosocial behavior, in particular the caring of the young, is at least as conspicuous as aggression in the animal world. In primates, many of the caregiving skills, far from being inborn, are learned, as shown by social isolation experiments; likewise mutual help and aggression are often learned, usually by imitation. And in all cases there are large individual differences. Both such differences and the learned nature of much of prosocial and antisocial behavior refute the sociobiological hypothesis that such behavior patterns are genetically determined and, moreover, have a single goal, namely to perpetuate the individual's genes. (See Zahn-Waxler, Cummings & Ianotti Eds. 1986, and Olweus, Block & Radke-Yarrow Eds. 1986.)

Now, some prosocial actions are socially disvaluable (e.g. complicity with criminals), whereas some antisocial actions are socially valuable (e.g. rebellion against oppression). This shows that we cannot equate "socially valuable" with "prosocial behavior". It will take us some time to arrive at the notion of social good (or good society), which will be introduced in Ch. 2, Sect. 1.4. Before tackling this problem we must turn from individual behavior to the behavior or working of entire social systems, such as families, gangs, firms, and states.

Whereas all biovalues and psychovalues are individual, some social values are individual while others are social or systemic, in the sense that they are attributed to social systems. Of course the evaluators themselves are individuals: social groups lack the brains necessary to perform evaluations. But sometimes the evaluators appraise entire social systems. For example, on the one hand productive work is an individual activity even when performed by a team, and it may be valuable for society as well as for the individual. On the other hand a public transit system, a hospital, or a school, has needs (hence values) and meets needs (hence values) as a whole. Thus, a public transit system needs vehicles and energy, as well as a work force and a management, and it can serve the members of a community only if its human components do competently their assigned jobs in a coordinated manner, and if the users of the system cooperate. The preceding can be compressed into

DEFINITION 1.11 Any bit of social behavior, and any social group, that preserves the integrity of a society, is *valuable* to the latter.

This is nothing but the social parallel to the principle that whatever contributes to the survival of an organism is valuable to the latter. ("Parallel" and not "identical" or even "a particular case" because societies are not alive.) Note however that what is valuable to a society as a whole may not be valuable to all of its members. For example, a criminal youth detention center is valuable to society as a whole even if it is not highly valued by its inmates. And a dictatorial government, though valuable to a theocratic, military, political, economic or bureaucratic caste, is disvaluable to most of its subjects. The crucial question at this point is not so much that of the functionalist anthropologists and sociologists, namely 'What promotes the preservation of society?'. The key question is whether a given social system is valuable to its members and to mankind as a whole. However, before answering this question we must insist on the systemic nature of whatever is valuable to a social group as a whole.

What is valuable to a social group as such cannot be distributed among either the members of the group or those who benefit from it: like public parks and public utilities, it is *indivisible*. Still, even supraindividual values are such because there are individuals who perform such evaluations on the strength of the benefits or burdens they receive from the social system in question. In other words, all social values are rooted to individual biovalues and psychovalues. Social systems, from the family to the international organization, are built, kept or reformed to meet what are ultimately individual needs or wants, legitimate or illegitimate. When they cease to meet such needs or wants they cease being valuable; and when this happens the time is ripe to reform or dismantle them. The view that social wholes are self-serving, and that their value hovers above the values cherished by its individual components, is a holistic myth that serves only to entrench privileges or to perpetuate obsolete social organizations. To help avoid this holistic trap we hasten to propose

POSTULATE 1.8 (i) A social group is socially valuable only if it helps its members attain or retain their good health and meet their psychologically legitimate wants. (ii) The most valuable of two social systems of the same type (e.g. firms or governments) is the one which is the more socially valuable to the more inclusive social group — i.e. the one that serves best the greatest number of people.

This assumption has two immediate consequences:

COROLLARY 1.2 The most valuable of all the (actual or possible) social systems in any given society is the one that serves best all of the members of the society.

COROLLARY 1.3 The most valuable of all the human social systems is the world system, i.e. the one that encompasses the whole of humankind.

The world system was only born five centuries ago, when all of the major human groups started to coalesce into a single one by way of conquest, colonization, and trade. The world system has needs of its own which transcend the national and regional ones, such as protecting the atmosphere and the seas from industrial pollution, halting the quick depredation of non-renewable resources, and preventing further international conflicts. Such global needs give rise to global or supranational values, such as a clean biosphere, a rational management of the world's non-renewable resources, and world peace. In turn, the fulfilling of such values, i.e. the attainment of such goals, poses technical and political problems on a world scale, that can only be tackled in a global manner. That is, the world system needs, for its own preservation, a global management in the interests of the greater number of individuals and their offspring. However, we must postpone the consideration of this tantalizing matter to Ch. 11.

Before drawing any further consequences from Postulate 1.8 we need another hypothesis:

POSTULATE 1.9 Only valuable items are worth the effort required to attain or retain them.

This proposition seems so self-evident as not to need explicit statement. Yet, actually we often engage in worthless actions, or neglect the important for the urgent, because we perform or receive the wrong evaluations. In any event, Postulates 1.8 and 1.9 jointly entail

THEOREM 1.1 A social system that fails to satisfy Postulate 1.8 is not worth keeping.

This is a reminder that social systems must be reviewed once in a while to ascertain whether they actually continue to serve (if they ever did serve) the (good or bad) purposes for which they were instituted. A case of obsolescence is that of a number of social programs originally launched in the US to correct some social injustices, but which in practice have failed to do so, in spite of which they are kept either because of bureaucratic inertia or political convenience. (See e.g. Gilbert, Light & Mosteller 1975, Mosteller 1981.) Actually the entire so-called welfare state — more properly called 'relief state' — is currently under scrutiny, not only because it hurts the wealthy but also because it has shown itself to fall short of its goal. In fact, while it relieves some misery it does not fully eradicate poverty, and it is incapable of satisfying two of the most basic human needs: those of having a useful and gainful occupation, and of feeling self-esteem. But the matter of basic human needs, hence values, deserves a new section.

2.4 *Summary*

According to axiological nihilism (e.g. physicalism and existentialism) there are no values except as fictions. On the other hand according to axiological idealism (e.g. Platonism and phenomenology) values are self-existent, hence absolute, ideas; as such they may but need not be "embodied" in concrete things. And according to theology-oriented axiologies, God is the source of all values.

In our own view there are values or, rather, items evaluated (positively or negatively) by animals, in particular humans. Animals value whatever they need or want. An animal may set store in an item either because it is necessary for its survival, its psychological well-being, or its functioning in a social group.

In other words, values can be biological, psychological, or social. For instance, wholesome food is a biovalue, company a psychovalue, and peace a sociovalue. But of course these values are interrelated rather than mutually independent. Thus access to food may depend on social standing, and the conservation or alteration of a social structure depends on the biological and psychological welfare of the individuals composing the society.

3. NEEDS, WANTS, AND VALUES

3.1 *Basic Needs and Legitimate Wants*

So far we have dealt with biovalues and psychovalues in general. Now we shall focus on basic human needs and legitimate wants as value sources. Moreover in this section we shall restrict our interest to personal needs and wants. The reason for discussing personal human needs and wants is that, in our view, these are the sources of human values.

The basic human needs can be grouped into three categories: physical or visceral, mental, and social. Table 1.2 exhibits a few items in each of these three categories. The reader may find it necessary to add or remove some items, perhaps with the help of some of the very few serious studies conducted on this crucial problem. (See e.g. Malinowski 1944, International Labour Office 1976, Ghai *et al.* 1977, McHale & McHale 1977, Streeten & Burki 1978, Harris 1979.) But before looking at any such sources he should be warned that it is far easier to investigate what people want, or rather *say* they want — e.g. by circulating questionnaires about individual preferences — than to determine their objective basic needs, e.g. in the fields of nutrition, health care, or education. As a consequence our objective knowledge of subjective wants (psychovalues) is far better than that of our objective needs (biovalues and sociovalues).

We shall distinguish two levels or degrees of need, primary and secondary, and shall define the corresponding concepts in terms of that of deficit or deficiency, i.e. whatever is lacking to achieve optimal survivorship:

DEFINITION 1.12 Let x be a biological, psychological or social deficit of a human being b in circumstance c. We call x

TABLE 1.2. Some basic human needs

Rank	Physical or visceral (β)	Mental (ψ)	Social (σ)
1	Clean air & water	Being loved	Peace
2	Adequate food	Loving	Company
3	Shelter & Clothing	Feeling needed	Mutual help
4	Safety	Meeting legitimate wants	Work
5	Physical activity	Learning	Participation
6	Health care	Stimulation	Mobility
7	Leisure time	Recreation	Social security

(i) a *primary need* of b in c if and only if meeting x is necessary for b to stay alive under c in any society;

(ii) a *secondary need* of b under c if and only if meeting x is necessary for b to keep or regain health under c in b's particular society;

(iii) a *basic need* if and only if x is a primary or a secondary need.

Examples. An adult human being needs about 2,500 kilocalories and 30 g proteins per day, a minimum of 5 m^2 of housing space, and 6 years of schooling, as well as the company of freely chosen fellow human beings, to live normally in a modern society.

As for the concept of a legitimate want or aspiration, in contrast to that of a mere fancy or whim, we propose to elucidate it through

DEFINITION 1.13 Let x be a psychologically legitimate want or desire of a human being b in circumstance c (Definition 1.7). We stipulate that x is a *legitimate want* (or *desire* or *aspiration*) of b in circumstance c and society d if and only if x can be met in d (i) without hindering the satisfaction of any basic need of any other member of d, and (ii) without endangering the integrity of any valuable subsystem of d, much less that of d as a whole.

Put negatively: An aspiration that can only be fulfilled at the expense of other people's needs, or the integrity of a valuable social group, is illegitimate in that group — even though it may be legitimate in another, perhaps more prosperous, group. Obvious examples of legitimate aspirations are to live close to one's beloved, enjoying nature and culture, and improving one's lot; of illegitimate aspirations: idleness, wanton destruction of lives or goods, and domination for its own sake. Just as living in society puts constraints on individual behavior, so individual needs and legitimate wants constrain the kinds of desirable society.

Both wants and needs are objective in the sense that they are deficits the elimination of which alters the state of the organism in certain circumstances. Hence both can be ascertained by scientific means. But whereas the satisfaction of primary needs is a matter of life and death, and that of secondary needs a matter of health or sickness, the satisfaction of legitimate wants is one of happiness or unhappiness. For example, food is a primary need, medical care is a secondary need, and entertainment a want.

3.2 *Values*

We are finally ready to define the concept, or rather concepts, of human values. We shall do so in terms of a human being's needs and wants, which in turn may be taken to be definable and measurable in the natural and social sciences. In fact we propose

DEFINITION 1.14 Let x be a thing, a property of a thing, or a process in a thing (in particular a mental process in someone's brain). We attribute x

(i) a *primary value* for human beings in circumstance $c =_{df} x$ contributes to satisfying at least one primary need of any humans, in any society, when in circumstance c;

(ii) a *secondary value* for human beings in circumstance $c =_{df} x$ contributes to meeting at least one of the secondary needs of humans under c in their particular society;

(iii) a *tertiary value* for human beings in circumstance $c =_{df} x$ contributes to meeting at least one of the legitimate wants (or desires or aspirations) of humans in circumstance c;

(iv) a *quaternary value* for human beings in circumstance $c =_{df} x$ contributes to meeting a fancy, i.e. a desire that is not a legitimate want;

(v) a *basic value* $=_{df} x$ has either a primary or a secondary value.

If we do not care for such analysis, or do not need it in certain contexts, we may lump all four value ranks into one, and speak simply of *good* items for humans in certain circumstances; i.e. we may use

DEFINITION 1.15 For any object x, x is *good* for human beings in circumstance $c =_{df} x$ has a primary, secondary, tertiary, or quaternary value.

However, we must not lose sight of the fact that tertiary and quaternary values are not universal, whence something good for someone (for realizing a ternary or quaternary value of hers) may be bad for someone else (e.g. the person from whom the object in question was taken). And we cannot take a further step and introduce the predicate "good" without qualifications, for nothing is good in itself, i.e. regardless of any evaluating subjects and in all circumstances. For example, there was nothing good or bad in the universe before the first animals emerged.

Note the following points. Firstly, values are not things, properties of things, or changes in things: instead, they are properties we, or other animals, attribute to the former. (Recall Sect. 1.1.) Therefore the publicity that calls low-priced commodities 'good values' is conceptually wrong: it should refer instead to 'valuable goods at low prices'.

Secondly, we can use Definition 1.14 to distinguish among primary, secondary, tertiary, and quaternary goods or value bearers, such as loaves or bread and yachts. In particular, we would define a *basic good* as one capable of satisfying a basic need, or realizing some primary or secondary value. This concept differs from Rawls's now classical one by the same name (Rawls 1971 p. 62). Indeed according to Rawls a primary good is one that "every rational man is presumed to want". We do not adopt this definition because (*a*) it is not accompanied by that of a "rational man" (by which Rawls does not mean the egoist who only cares about maximizing his own expected utilities), and (*b*) it presupposes an economy of plenty, for in one of subsistence or scarcity a rational person, and a fortiori a rational and moral one, will not want what he does not need, for he will not defy the envy and ire of his neighbors. In our view any goods which are not basic are secondary, even if they are capable of satisfying legitimate aspirations, because they may or must be dispensed with in the event of scarcity.

Thirdly, Definition 1.14 would be attacked by the intuitionists (e.g. Moore 1903) as a "naturalistic fallacy". But this only goes to show that axiological intuitionists overlook the factual root of values and morals — whence they can be of no help in ordinary life, much less in public life. Basing values on needs and wants, instead of removing them from the world, has two advantages. Firstly, it builds a bridge between value theory and ontology. Secondly, it entails the possibility of basing value judgments on scientific findings about needs and wants, hence of ultimately turning axiology and ethics into scientific disciplines. (See e.g. Bunge 1960, 1961, 1962a.)

However, there is one *prima facie* counterexample to the identification of value with need or want satisfaction, namely no less than life itself. Indeed, one may well ask what need or want does life satisfy. Put negatively: Why should we cherish life, let alone regard it as inviolable or even sacred, if living satisfies no obvious need but, on the contrary, is the source of all needs and wants? We submit that this puzzle originates in an individualistic perspective, i.e. one in which the individual is seen as a member of a loose aggregate rather than a component of a tightly knit system. The puzzle evaporates in a systemic perspective, where every human being is regarded as (actually or potentially) playing some role in a society, and thus being able to satisfy some of the needs and wants of other people. For example, a baby, even if severely handicapped, satisfies his mother's need for giving care and being loved; a worker satisfies a need of the management, and the latter meets some of the former's needs, and both together are supposed to meet needs or wants of some consumers; a teacher meets the learning needs of her students, who in turn satisfy the former's need and wish to teach — and so on. In short, although life in the abstract does not derive from any need, we value every living person for contributing (actually or potentially) to other human lives. Besides, every one of us came to life as a result of physical, emotional or social needs or wants of our parents. And we normally enjoy satisfying our own needs and wants as well as some of those of others — and this is what life is all about.

3.3 *Value Categories and Orders*

The basic human needs and the corresponding values do not come in amorphous collections but are partially ordered within every category. For example, clean air and water precede nutrition, which in turn precedes shelter and clothing; being loved dominates loving, which in turn dominates feeling needed; and peace is more important than company, which in turn is more important than mobility. (For alternative hierarchies see Rogers 1964 and Maslow 1968.)

In short, $\mathscr{C}_i = \langle V_i, \succsim_i \rangle$ for $i = \beta, \psi, \sigma$, where $\mathscr{C}_i$ is a value category and V_i the corresponding set of particular values. The main point of grouping values into categories is to avoid comparing values out of context. For example, it would be mistaken to state that truth is (absolutely) more (or less) important than friendship or beauty, or that reasearch is more (or less) important than physical exercise. Value judgments make sense only in context, and when in context some of them are objectively true. For example, whereas "Truth is more important than beauty" is true with reference to science, it is false relative to art.

It would be mistaken to try and order the value categories $\mathscr{C}_\beta$, $\mathscr{C}_\psi$ and $\mathscr{C}_\sigma$ themselves, for most visceral and mental needs cannot be met except in society, i.e. in the company and with the help of others. Primates and other higher animals are biopsychosocial things, whence they have biovalues, psychovalues and sociovalues. The importance of the latter cannot be exaggerated. No human individual is an isolated and self-reliant atom: every one of us, even the most miserable outcast and the humblest hermit, belongs, however ephemerally, to at least one social group. And meeting a basic need or a legitimate want in any of the three categories calls for satisfying some basic needs in the other two. For example, there is no learning on an empty stomach and in a social vacuum. Because of such interdependencies among the three basic value categories, these do not form a hierarchy but a triangle: See Figure 1.2(*b*).

We compress the preceding into

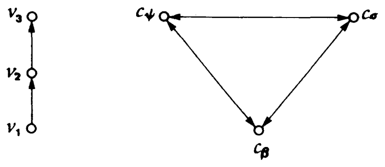
POSTULATE 1.10 (i) There are three categories of basic human needs and legitimate wants: physical or visceral ($\mathscr{C}_\beta$), mental ($\mathscr{C}_\psi$), and social ($\mathscr{C}_\sigma$). (ii) For every such need or want there is a corresponding value, viz., that of meeting that need or want. (iii) Every one of the three value categories is ordered by a relation $\gtrsim$ of dominance, i.e. it constitutes a partially ordered set. (iv) The three value categories are interdependent, forming a triangular loop: i.e., every value in one of the categories can only be realized if some values in the other two categories are realized as well.

The following are immediate consequences of the preceding axiom:

COROLLARY 1.4 It is impossible to order all of the human values in a single linear order or hierarchy (i.e., there is no single "table of values").

COROLLARY 1.5 No value is absolutely preferable to any other.

It may be retorted that human survival is the basic or absolute value for human beings. Indeed it is, but not in the sense of merely visceral survival, without mental life, for such "vegetative" life cannot be



(a) (b)

Fig. 1.2. (*a*) Within every value category (visceral, mental, or social) the values form a linear sequence or hierarchy, in which each either dominates or is dominated by others. (*b*) The three categories constitute a triangular loop. Every node represents a collection of values of a kind: visceral ($\mathscr{C}_\beta$), mental ($\mathscr{C}_\psi$), or social ($\mathscr{C}_\sigma$). Every double arrow represents the mutual dependence of such value groups.

enjoyed by anyone; nor in the sense of survival out of society, for nobody can survive without the help of others, and nobody can enjoy life without the company of peers. The maximal value is not personal survival but *full (visceral and mental) survival in society* — or, rather, optimal well-being (physical, mental and social) in a society worth living in. However, the explicit postulation of this maxim will have to await the elucidation of the concept of well-being, to be tackled in the next chapter.

3.4 *Summary*

Values emerge from animal needs or wants. Humans and other higher gregarious animals have requirements of three types: physical or visceral, mental, and social. These are interrelated. For example, a person cannot lead a normal social life unless she is well nourished and mentally balanced. In other words, needs and wants come in clusters.

We group needs and wants into four clusters: primary needs (to be met for sheer survival), secondary needs (to be satisfied in order to enjoy good health), legitimate desires (to be met for the sake of reasonable happiness), and fancies. Every one of these clusters contains biological, psychological, and social components.

Every cluster of needs or wants is the root of a group of values: primary, secondary, tertiary, and quaternary. Only the first three are genuine or legitimate. See Figure 1.3. This categorization will have momentous consequences for our views on rights and duties, i.e. our ethical doctrine.

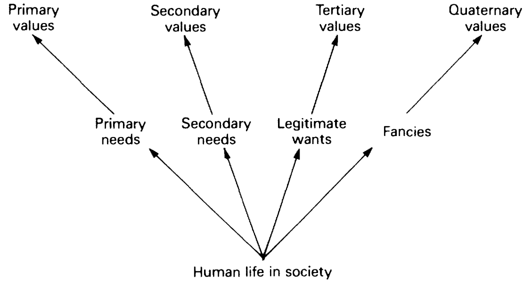


Fig. 1.3. Living in society involves meeting basic needs and certain wants, which in turn makes us value whatever is capable of satisfying these needs or wants. The realization of primary values ensures sheer survival, that of primary and secondary needs guarantees well-being, that of primary, secondary and tertiary, reasonable happiness. Mere fancies or whims are the root of artificial values.

CHAPTER 2

WELFARE

All animals pursue their own well-being, either automatically or deliberately. Those who do not perish. Human beings, particularly since the rise of civilization, seem to be the exception. They are the only animals to have invented devices designed to prevent some conspecifics from attaining well-being or even from staying alive: slavery and serfdom, jails and gallows, war and conquest, myth and propaganda, and much more. Hence the importance of any moral and legal codes enshrining the right to the pursuit of well-being.

The principle that all humans, or at least all free humans, have the right to pursue their own well-being, is ancient. Its explicit formulation goes back to at least Aristotle's *Nichomachean Ethics*, rightly called a *eudaimonistic* ethics for holding that the goal of man is *eudaimonia* or well-being. This idea was alien to Buddhism and it was opposed by the early Christians and later on by the protestants (particularly Luther and Calvin). In modern times it was derided by Kant and, later on, by fascists and nazis. To all of them, though for different reasons, human life is suffering and struggling for transcendental goals, such as nirvana, the after-life, the greater glory of God, or that of the omnipotent State.

The American Declaration of Independence (1776), drafted by Jefferson, stated that all (white male) men have "inalienable rights", among them "life, liberty, and the pursuit of happiness". Although this document tacitly excluded slaves from such rights and did not give women any political rights, it was a humanist manifesto: it spoke of life, not after-life; of liberty, not submission; and of happiness, not the cross. Its descendant, the more influential *Déclaration des droits de l'homme et du citoyen* (1789), enshrined similar basic rights although it made no reference to happiness. The liberals and socialists continued and strengthened this tradition, working for social reforms, or even revolutions, aiming at securing the well-being or even happiness of everyone.

Since so much has been written and legislated about well-being and happiness over the past two centuries, one would think that these concepts are quite clear. Not so: the notions of well-being and happiness, like those of fairness and justice, freedom and democracy, are still inexact. One of the reasons for this may be that they have remained for the most part in the foggy regions of ordinary knowledge and ideology. One of our tasks will be to try and elucidate those notions, starting with well-being and happiness.

1. WELFARE AND HAPPINESS

1.1 *From Bentham to Decision Theory*

The first modern treatise on happiness was Bentham's (1789). He equated happiness with "enjoyment of pleasures, security from pain", and listed fourteen "simple pleasures" (including those of benevolence and imagination) and twelve "simple pains", among them those of privation and enmity (Ch. V). Moreover Bentham popularized the Utilitarian Principle, of the greater happiness of the greatest number — actually first proposed by Helvétius and adopted by Priestley. Later utilitarians, such as Mill (1863), Moore (1903) and Sidgwick (1907), spoke of good or utility rather than happiness. But in every case the key notion remained vague and also stray, i.e. outside any theory (hypothetico-deductive system). Moreover Moore insisted that "good" is undefinable and can only be intuited. Actually we are dealing here with a whole *family* of vague concepts — those of happiness, pleasure, good, and utility — with a common core.

The first serious attempt to elucidate the notion of utility (or subjective value) was that of the decision theorists, who were motivated by psychology and social science rather than by ethics. (See von Neumann & Morgenstern 1947, Luce & Raiffa 1957.) However, it is now clear that theirs was a glorious failure. Glorious because they attempted a grand enterprise — that of building a single and simple foundation for psychology and all of the social sciences — and a failure for a number of reasons.

Some of the reasons for the failure of decision theory and its offspring, game theory, were mentioned in Vol. 7, Ch. 5, Sect. 5.2. These theories are not scientific because the probabilities and utilities occurring in them are subjective, i.e. eyeballed rather than assigned by either calculation or measurement. Consequently, by conveniently assigning numerical values to probabilities and utilities, one may prove anything — e.g. that selfishness pays off, or that cooperation does. Further reasons for the failure of decision theory is that the utilities of different persons can be neither compared nor aggregated; in particular, it makes no sense to speak of the aggregate utility of something for an entire social group. (See e.g. Sen 1979.)

However, the root troubles with decision theory, game theory, and their ethical companion, namely utilitarianism, are even deeper: to wit, vagueness, irrefutability, and unconcern with human needs and wants, hence with genuine values. Let us take a quick look at these flaws. The first defect is obvious: Decision theory does not tell us exactly *what* utility is. Sometimes 'utility' is read as value, at other times as welfare, and more often as profit or payoff — particularly when proposing payoff matrices pulled out of a hat. But these are quite distinct concepts. For example, murder and exploitation may be profitable but they do not improve the welfare of their victims; on the other hand a sacrifice may have a disutility to its doer but not to its beneficiary — and, since such positive and negative utilities are neither comparable nor additive, how are we to decide whether the total utility is greater than zero, hence the sacrifice worth performing?

The methodological trouble with the Principle of Utility is that it is hardly refutable, for it covers every instance of deliberate human behavior, including any attempt to refute utilitarianism. Indeed, the principle does not state any relations between particular actions, or the need for them, and the associated outcomes and their utilities. Consequently if a person performs now a certain action, next another action with the opposite outcome, and finally no action at all, utilitarians ask us to believe that in all three cases the person has acted so as to maximize her expected utility. (Example: Person *A* maximizes her expected utility by overachieving, listening to classical music, and loving; person *B*, by underachieving, listening to hard rock, and hating; person *C* by enjoying life, and person *D* by retiring to a convent or even committing suicide.) In sum, according to utilitarianism people maximize their utilities *whatever they do* deliberately. The Principle functions very much like a tautology: it holds come what may. If it be rejoined that the theory does not attempt to describe actual behavior but to prescribe rational action, then it must be said that (*a*) preferences (psychovalues) are not rational pointers, and (*b*) individual utility maximization is usually sheer selfishness, hence morally objectionable. (More on the latter in Ch. 7.)

The third flaw of decision theory and utilitarianism listed above is their unconcern with basic needs and legitimate wants. All they are interested in is whether the outcomes of the actions in question have consequences that their agents value positively for some reason or none. In those theories there is no room for the very sources of genuine value (Definition 1.14). Therefore we cannot use them to build a correct value theory. However, we shall see in Ch. 7, Sect. 3.3 that utilitarianism does have some merit in ethics and in action theory after all. But now let us go back to the concept of well-being.

1.2 *Definitions*

We propose the following

DEFINITION 2.1 An animal (in particular a human being) is in a state of *well-being* (or *welfare*) if and only if it has met all of its basic needs (Definition 1.12).

As emphasized in Ch. 1, humans are biopsychosocial entities. Hence they have needs, and therefore values, of three different though mutually related categories: Recall Figure 1.2. Consequently there are three distinct though mutually related sources of well-being — physical, mental, and social — and just as many as sources of discomfort. Well-being is a vector, not a scalar. Therefore it is mistaken to try to reduce well-being to the meeting of a single need, e.g. food, sex, freedom, or money. All unifactorial (or monocausal) views of well-being are then false, even though some of them contain a grain of truth.

Nor is well-being a state to be attained or retained passively: spoon-feeding is only for babies and the infirm. All healthy animals are active, hence they do not feel well unless they can move about and, in the case of humans, do something useful. Well-being then is a matter of doing as well as being, and of giving as well as taking. Although Aristotle made this point more than twenty-three centuries ago, and although psychologists and anthropologists have amply confirmed it, we must emphasize it in view of the assumption of conservative social scientists (or rather ideologists) and moral philosophers, that with few exceptions we are basically lazy and selfish, and thus ready to live on public welfare (as "free-loaders") rather than earning our upkeep. What is true is that undernourished, chronically unemployed and marginal people end up by becoming disheartened and lack the energy and the will to fight a losing battle.

In principle the concept of well-being can be further honed with the help of the state space concept (for which see e.g. Vol. 4, Ch. 1, Sect. 2.2). Take the totality of N known independent variables that make up the visceral, mental and social state of a human being. Every one of these variables may be conceptualized as an axis in an N-dimensional abstract space. The various laws and constraints that interrelate or restrict these variables define an N-dimensional box L (for "live"). Every point within L represents a state s of the animal. Now, for an animal to be in a state of (visceral, mental and social) well-being, every one of the N variables must range over a comparatively narrow interval. These N intervals define a much smaller box W included in L: this is the *well-being box*. In other words, a state of well-being is representable as an arrow s with origin in the origin of the N-dimensional coordinate system, and tip lying inside the W box: see Figure 2.1. Some of the components (projections) of the well-being arrow will be large whereas others will be small, depending on the scale, the biospecies, and even the individual. Thus a person may be healthy but joyless, and another sick but happy.

Besides being exactifiable, well-being is measurable, at least in principle. In the case of humans there is a whole battery of well-being indicators, from calorie and protein intake to fraction of leisure time devoted to voluntary work. The most telling of all such well-being indicators is perhaps longevity, which ranges between 37 (Afghanistan) and 77 (Sweden).

The momentary state of well-being, or rather discomfort, can also be measured in the laboratory. An animal subjected to extreme crowding or harassment is under a stress that, if prolonged, can cause remarkable alterations in the size and shape of certain organs (e.g. the thymus), as

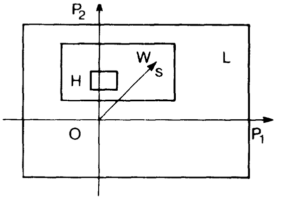


Fig. 2.1. State space of an organism of a given species in a given environment, oversimplified to two axes, every one of which represents a property of the organism in its environment. The large box L contains all the points representing the possible states of the living organism, whether healthy or sick, happy or unhappy. The smaller box W represents possible states of welfare. The tiny box H represents the possible states of happiness.

well as an overproduction of corticosterone (cortisol in the case of humans). For this reason the concentration of corticosterone (or cortisol) in blood is used as an objective indicator of stress, hence of discomfort, in the higher vertebrates. Unwell-being, or discomfort, has become a subject of scientific research in social science, psychology, psychoneuroendocrinology, and psychosomatic medicine, as well as in applied ethology with reference to domestic animals. (See e.g. Zayan 1984 and Zayan Ed. 1985.) Since farm animals are not in the habit of filling out psychosociological questionnaires, but can be subjected to objective tests, we probably know more about their welfare conditions than about our own.

The concept of happiness is much narrower than that of well-being. In a state of well-being all of the basic needs (or primary biological, psychological and social drives) have been satisfied. Yet an animal may still be dissatisfied because some of its wants have not been met or, worse, it has no expectations that they will be met. Think of zoo animals or of prison inmates. Happiness is a particularly satisfactory state of well-being: it is attained when the animal is satisfied with having achieved certain goals or with the prospects for attaining them. More precisely, we stipulate

DEFINITION 2.2 An animal feels *completely happy* if and only if it believes to have the ability and opportunity of meeting all of its needs and wants.

Like well-being, happiness is a state (or rather a set of states) that an animal can be in: See Figure 2.1, box *H*. This fact has important consequences. One of them is that there can be no happy *societies*, but only societies some members of which can be happy to some extent. A second consequence of the fact that well-being and happiness are states, not things, is that they cannot be *aggregated*. I.e., it makes no sense to try to add up the degrees of well-being, or of happiness, of the various members of a social group, in order to obtain the total amount of well-being or of happiness in the group. (The tacit assumption of the classical utilitarians, that such an aggregation is possible, is just as mistaken as the idea that population densities or temperatures are additive.) A third consequence of the fact that well-being, and in particular happiness, are states of individuals, is that they cannot be *distributed* anymore than an intention or a smile can. Only some well-being and happiness *means*, such as food and caregiving, can be distributed among people.

Note that, unlike well-being, which is an objective biopsychosocial state, happiness is a transient feeling — a subjective experience. Being subjective, the conditions for attaining it vary from person to person. Thus a person may have everything she really needs and wants, yet she may feel unhappy — e.g. because she fears to lose her possessions — whereas another individual, blessed by a sanguine temperament, may have moments of happiness even under duress.

Happiness is, in short, not only a matter of objective circumstances, in particular material means, but also of mental set or disposition. (But of course the latter is a child of circumstance as well as of genetic make-up.) Hence, while a physician may find nothing wrong with a patient, a psychologist may discover unhappiness, derived, for example, from an unjustified phobia. Accordingly the (scientific) psychologist may teach her patient to remove some of the obstacles to her (temporary) happiness. Shorter: We can learn to become happy — e.g. by counting our blessings every time we suffer a mishap.

As a matter of fact some philosophers too have had something to say about happiness. For example, the ancient Stoics recommended reducing one's needs and wants instead of multiplying them as we do in consumer societies. (In view of the increasing scarcity of many of the raw materials employed in manufacturing industrial goods we ought to heed this advice. More on this in Ch. 11.) But other philosophers have blundered with regard to happiness, some for holding that it is unattainable except in the after-life, and others for equating it with goodness. The former view is untenable, for happiness, like any other feeling, cannot be known without experiencing it; and if we cannot experience it while alive, how could we experience it when dead? As for the identification of happiness with goodness, it is mistaken because goodness can be predicated of objects of many kinds, whereas happiness can only be predicated of the mental states of some animals. In particular, we may talk about good societies but not — except elliptically — about happy ones, for societies are mindless.

1.3 *Postulates*

It is a well known fact of life that there are mutually incompatible (or conflictive) desiderata. Moreover, for every desideratum there is a counterdesideratum, though not always one that it is rational to hold. For example, the single-minded pursuit of happiness is bound to cause pain to self and others; obsession with security endangers peace;

ownership causes worries; the advancement of anyone is likely to hurt others — and so on. This suggests

POSTULATE 2.1 The full realization of any given value is incompatible with that of some other values.

The conjunction of this assumption with Definition 2.2 entails

THEOREM 2.1 Nobody can be completely happy.

This result refutes the hedonistic thesis that the goal of man can be, is, or ought to be the attainment of full happiness. Knowing that this goal is unattainable, the rational being will settle for something less, namely reasonable happiness, as defined by

DEFINITION 2.3 An individual is *reasonably happy* if and only if she is

(i) in a state of well-being, and

(ii) free to pursue her legitimate wants.

Even reasonable happiness is largely a subjective — mainly emotional — state and, as such, elusive and short-lasting. Though attainable, reasonable happiness is actually attained by less than one person out of ten in the whole world. (Remember that most people live in Third World countries, where poverty is rampant, one out of two people do not have safe water to drink, and one out of five do not get enough food for an active working or playing life: see e.g. Sivard 1987.) When living at a subsistence level or below there are few opportunities for happiness. Most of the people in the present world do not aim for maximal happiness: They feel satisfied if they stay alive and give their dependents a chance to survive. Rather than dreaming a calculus of felicity they dream of stumbling on a survival martingale.

Yet in all societies we see plenty of examples of absurd evaluations, some leading to what Veblen (1899) called 'conspicuous consumption' (engaged in for the sake of appearances), others to the attainment of transient happiness to the detriment of health. Obvious examples are the preference for drugs, such as alcohol, over food; of cars over education; and of entertainment over health care. This all too common irrational evaluation pattern calls for the adoption of

NORM 2.1 Long-term well-being and, a fortiori, reasonable happiness, calls for the following ranking: Meeting primary needs (survival) ought to precede meeting secondary needs (health), which in turn should precede meeting legitimate wants, which ought to dominate the satisfaction of fancies.

Using Definition 1.14, this obvious yet often forgotten rule translates into: Primary values precede secondary values, which in turn precede tertiary values, which dominate quaternary values.

Finally, note that the preceding principle is a norm or rule not a postulate or basic assumption. The reason is that people do not always observe it.

1.4 *Social Welfare*

The pursuit of well-being, and even more so that of reasonable happiness, cannot be a "solo" endeavor: it can only be conducted in society, and moreover in a society with special characteristics. There are several obvious reasons for this. Firstly because congenial company is no less than a basic need: normal people do not feel well unless they sense that they "belong" in at least one social group, usually more than one. Only extremely abnormal people, such as psychopaths and autistic individuals, imagine themselves to be isolated atoms. But of course not even such psychological deviants are self-reliant: they depend on others for survival. Normal people seek the help, respect and affection of others, and attempt to reciprocate them: these are important components of the well-being vector. (This is why ostracism and solitary confinement are among the most cruel of all punishments.)

A second reason for linking the pursuit of well-being to society is that everyone, even the most self-reliant individualist among us, needs the help of others to meet her basic needs and legitimate wants. (Robinson Crusoe would not have survived his shipwreck without the skills he had learned from others, as well as the tools he salvaged from his ship.) A third reason that personal well-being and reasonable happiness depend very much upon social organization is that mutual help, though necessary, is not sufficient. We also need social organizations, both public and private, facilitating mutual help or cooperation, dampening competition, and supplying such public services as security, sanitation, education, and transport. Not even anarchists dare deny this.

Now, not all societies can give everyone the chance to pursue her well-being, much less her (reasonable) happiness. Only the right combination of economic and cultural riches with social (distributive) justice can do that trick. The elucidation of this idea calls for three definitions, starting with

DEFINITION 2.4 A society is

(i) *miserable* (M) if and only if it lacks the (economic, cultural and political) resources required to meet the basic needs of all of its members — whence it cannot secure the well-being of them all;

(ii) *poor* (P) if and only if it has the (economic, cultural and political) resources required to meet the basic needs but not the legitimate wants of all of its members — whence it can secure the well-being but not the reasonable happiness of all of them;

(iii) *properous, rich, or affluent* (R) if and only if it has the (economic, cultural and political) resources required to meet the legitimate wants as well as the basic needs of all of its members — whence it can secure not just the well-being but also the reasonable happiness of them all.

In other words, in the miserable societies — all those in the Fourth World and many in the Third — few if any can realize the primary and secondary human values. In the poor (or subsistence) societies — most of those in the Third World — there is just enough for nearly everyone to realize the primary and secondary human values. In the rich societies — most of those in the advanced industrialized nations — everyone can realize the primary, secondary and tertiary human values. As for the quaternary values, which satisfy fancies or whims, no society could possibly realize them for all of its members, but nearly every contemporary society contains a handful of privileged people who do realize a good many of them — at the expense of the greatest number.

The above definition contains the modal expression 'can' (or its denial): it is about possibility, not actuality. This is because the resources in a poor or even a rich society may be distributed so unevenly that in fact only a privileged minority can expect to be in a state of well-being. (Think e.g. of the dozen families and the couple of transnational companies that own most of the good land in El Salvador.) Therefore we also need

DEFINITION 2.5 A society has a *just* social structure (J) if every member of the society can attain well-being, or even reasonably happiness, without others suffering from it. Otherwise the social structure is *unjust* ($\bar{J}$). In a just society the well-being of everyone depends on that of the others. (Such a society would seem not to satisfy Pareto "optimality" — the see-saw condition where no one can be better off without making someone worse off. However, Pareto "optimality", a sacred cow of mainstream economics and political science, is satisfied by *any* partition of the social cake, in particular by an extremely inequitable one: See Moessinger 1989. Hence there is no reason to regard it as a desideratum.) In a just society the improvement of anyone's lot parallels that of all others by virtue of cooperative mechanisms. We shall return to this matter in detail in Ch. 11.

The previous definitions have introduced three predicates (M, P, R) concerning social (economic, cultural and political) resources, and one predicate (J) concerning the distribution of such resources. Obviously, there in principle six possible combinations of these predicates:

MJ = Miserable & Just , $M\bar{J}$ = Miserable & Unjust
PJ = Poor & Just , $P\bar{J}$ = Poor & Unjust
RJ = Rich & Just , $R\bar{J}$ = Rich and Unjust

In MJ societies everybody is under the subsistence level. In $M\bar{J}$ societies everybody is miserable except for a handful of privileged families. In PJ societies everybody attains a state of well-being but nobody can satisfy any legitimate aspirations. In $P\bar{J}$ societies some are reasonably happy at the expense of he discomfort or suffering of the greatest number. Ditto in the $R\bar{J}$ societies, except that in these the fraction of people who can attain reasonable happiness (namely those forming the middle class) is far larger than in the poor societies. Only in RJ societies can everyone become reasonably happy.

The states PJ and RJ are so desirable, and so within reach of most societies, that they deserve names of their own:

DEFINITION 2.6 A society is said to be

(i) a *welfare society* if and only if it is poor but just, so that everyone is assured a state of well-being;

(ii) a *good society* if and only if it is just as well as rich, so that everyone has the opportunity of becoming reasonably happy.

Note in the first place that prosperity together with social (distributive) justice guarantees well-being for all but not happiness for all, since happiness is an emotional state that a few people are constitutionally incapable of enjoying.

Secondly, we have defined the concept of a welfare *society* not that of welfare *state* or government. Indeed, in a welfare society, and a fortiori in a good one, there are no needy people except as a result of natural catastrophes, so that there is no need for the state to take care of them on a regular basis. On the other hand the so-called welfare state attempts to correct some of the most glaring social inequalities by ministering to the basic needs of the destitute, though not to their legitimate wants — whence it is best called the *relief state*, or even the stop-gag or makeshift state. In a welfare society there are neither rich nor poor, whereas a relief state — characteristic of the most advanced societies of the $P\bar{J}$ and $R\bar{J}$ types — is a sort of Robin Hood that redistributes a small fraction of the excess wealth of a minority. We shall take a closer look at this matter in Ch. 11.

Finally we stick our neck out and make a value judgment:

POSTULATE 2.2 The six possible types of society are ordered as follows:

$$RJ > PJ > R\bar{J} > P\bar{J} > MJ > M\bar{J}$$

This is neither a strictly socioeconomic nor a purely moral judgment: it is mixed. According to it, neither prosperity nor social justice suffices by itself to build a welfare society, much less a good one: both factors are necessary. (Geometrical analog: the base and height of a rectangle.) Still, between two unjust societies we should prefer the richer one because it gives more people the chance of attaining well-being or even reasonable happiness.

Of course, ours is not the only possible evaluation of society types. A mainstream economist would only pay attention to the first member of every pair and disregard all matters of justice; he would propose the ranking: $R > P > M$. On the other hand a moral philosopher or political scientist might focus exclusively on the second member and state simply: $J > \bar{J}$. The former would then lump together just and unjust societies, and might argue that distributive justice is pointless where there is precious little to distribute. Moreover he might argue that, wherever there are surplusses, social injustice may motivate people to push for social changes aimed at a more just distribution. On the other hand the moral philosopher and the political scientist would lump together miserable, poor and rich societies, and might argue that prosperity is immoral unless shared by all. Further, they might argue that people in a just society, however poor, may be motivated to create the resources required to meet the basic human needs and legitimate wants. Since each of the two rival views has a grain of truth, we had better combine them. This is what we have done by proposing Postulate 2.2.

1.5 *Summary*

Animals are said to be in a state of well-being only if they have satisfied all of their primary and secondary needs. And humans may attain —

within bounds and over certain periods — reasonable happiness if, in addition to enjoying well-being, they can work to meet their legitimate wants. The latter are the desires that can be satisfied without seriously harming anyone else. A society where everyone enjoys well-being deserves being called a 'welfare society' — not to be mistaken for a relief society. And a society where everyone has a chance to attain reasonable happiness through his own effort and with the help of others may be said to be a good society.

Now, humans are not isolated atoms but components of social systems. Hence the actualization of values depends not only on personal ability and effort but also on the social structure, i.e. on the relations that hold people together. But most of the societies in existence do not afford most of their members the opportunity of pursuing their reasonable happiness. Just proclaiming the right to pursue it is not enough: the means for pursuing reasonable happiness must be available. To put it in terms of values: The value systems prevailing in most contemporary societies are not geared to giving all of their members the chance of meeting their basic needs, much less their legitimate wants. The much-needed social reforms will involve substantial changes in the prevailing value systems. But this is the subject of the next section.

2. VALUE CONFLICT AND CHANGE

2.1 *Value Conflict*

All humans, and possibly all animals, face value conflicts. These can be internal, interpersonal, or intergroup. An individual faces an *internal* value conflict when she is torn between mutually incompatible values, such as engaging in a pleasurable occupation and discharging a painful duty. Two individuals enter into an *interpersonal* value conflict either when they hold mutually incompatible values or when they hold the same values but the means available to realize these are scarce. And two social groups may enter into an *intergroup* value conflict when their respective value systems are at odds with one another or when they share the same values but compete for the same resources to realize them.

Value conflicts are so pervasive that is is difficult to imagine individuals and societies totally free from them. So, we may as well suggest

POSTULATE 2.3 All human beings and all social groups face value conflicts, some internal and others external (interpersonal or intergroup), arising either from holding mutually incompatible values or from a scarcity of means to realize the same values.

What can be done about value conflicts? Either nothing at all, fanning them, letting one of the rival values override its dual, or adopting a compromise. The course to be chosen depends on the nature of the conflict as well as on the agents's goals and means. Small conflicts must be viewed with tolerance and allowed to subsist. But those which are matters of life or death cannot be skirted. Yet even such serious value conflicts usually can be resolved, at least temporarily, by bargaining towards a compromise. For instance, every firm faces a conflict between profit and service; if its managers wish to stay in business they will not resolve such conflicts by maximizing either profit or service at the expense of the other, but will opt for a compromise involving submaximal profit and service.

We make bold and generalize:

POSTULATE 2.4 (i) Every value conflict can be resolved through some compromise whereby neither one of the values in question is either fully realized or fully sacrificed. (ii) Compromise is preferable to demise.

This axiom entails that nothing, not even safety and security, should be pursued to its maximum or minimum. For, if it were, then either the corresponding counterdesideratum or the people involved in the conflict would be sacrificed. For this reason, whereas in physics extremality principles (such as Fermat's and Hamilton's) hold sway, in human affairs optimal goals and courses of action lie usually between minima and maxima: satisfaction is the ticket, not maximization or minimization. (See March & Simon 1958, Beer 1966, Bunge 1982a, Afriat 1987.)

Most value conflicts are resolved through compromise, but no compromise is everlasting. A compromise may become obsolete either because of the defection of one of the parties, or because the value conflict that motivated it has ceased to exist or has been postponed by a more important one. For example, the ideological conflict between the two superpowers is increasingly seen as having been superseded by the survival-extinction dilemma. The old slogan "Better dead than red (or black)" is being replaced with the more reasonable "Better alive than dead". Only living people can argue and fight over economic and political issues. As we come nearer to the brink, some of us realize that the basic (primary and secondary) values take precedence over all others. A radical change in values is occurring worldwide under our noses — albeit, far too slowly.

Value changes and social evolution go hand in hand. As people adopt new lifestyles, some values disappear while others decay slowly and still others are actively fought against. But value changes are justified only if the old values are no longer rooted in basic needs or legitimate aspirations. For example, the young people who revolted against the "establishment" in the 1960's and 1970's were right in attacking international aggression and environmental degradation, but they were wrong in attacking science. Any global attack on a value system, without pausing to weigh every one of its components, results at best in a waste of energy, at worse in a backlash, and this for the simple reason that the value system of any viable society contains some values worth keeping — otherwise the society would not be viable. Besides, a rebellion against a value system is ineffective unless accompanied by the proposal of a new and superior value system. The reason is that man cannot live in a value vacuum any more that he can live in a social vacuum. In turn, the reason for such *horror vacui* is that values point to goals, and goals guide deliberate action. A person without a coherent value system is therefore like a leaf in a storm.

The search for alternative values starts with a rational criticism of the prevailing ones. We must begin by ascertaining whether the received values stem from basic needs or legitimate aspirations in our own society. (Recall Ch. 1, Sect. 2.4.) In particular, we must find out whether any of the currently held values originates from an illegitimate reversal of a means-end relation. Clear cases of such reversals are gluttony (living to eat instead of eating to live), professional sports (living for sports instead of the reverse), and the worst offenders: living only for pleasure, possessions, profit, or power, instead of seeking (harmless) pleasure, possessions, profit and power to secure survival. We submit that the unbridled pursuit of these four P's — pleasure, possessions, profit and power — with little or no consideration for other values, has driven mankind to the brink of extinction. More on this later on.

2.2 *Ideals*

People with only short term objectives are bound to go adrift. Only individuals with ideals can plan and build their own lives. But they will not succeed if their ideals are quixotic or if, being realistic, they do not join others with similar ideals. The yahoos who laugh at ideals, in particular at moral and social utopias, are condemned to live the lives other people have planned for them. We should not deride ideals but only unattainable ones: and we should oppose only bad ideals, such as those of living only for pleasure, possessions, profit, or power.

Ideals can be personal or social: they may concern individuals or social groups. In either case an ideal is a conceptual model or blueprint for a future state of a system, and moreover a state that someone desires for some moral or social reason. In the case of human individuals or collectivities, as well as of systems under human control — such as farms, factories, and schools — an ideal is a state that we seek the system to attain for someone's benefit. Such a goal-state is rarely attained in one shot: if attained at all, it is reached through a process composed of intermediate means-ends pairs, and involving a number of corrections called for by the realization of defects in the original plan or by unforeseen circumstances. The more complex the system, and the more ambitious the goal, the more numerous and tougher problems it poses, and so the vaster and deeper knowledge is required to steer the system to the goal-state with the least waste. (See e.g. Ackoff and Emery 1972, and Vol. 7, Ch. 5, Sect. 4.)

However knowledge, though necessary, is insufficient to attain any social ideals: the active participation of all concerned is required as well. Such participation will be forced if the ideal is not shared by the majority, and voluntary if it is shared. Now, a social reform is unlikely to be lasting unless it is beneficial to the greatest number and well designed. In other words, the key to successful social reform is a combination of sociotechnology with popular participation. Revolutionary enthusiasm without rational planning can only lead to bloodshed and disappointment; and expertly designed blueprints that fail to ignite popular involvement are bound to remain on the drawing board. And no popular adherence can be expected unless the project in question aims at realizing widely shared desiderata, even if these are base, such as those of revenge and conquest.

Ideals are attainable or not; if the former, they may or may not be attained, depending on our knowledge and the means at our disposal. Ideals such as immortality, omniscience, universal and everlasting plenty or idleness, and complete happiness, are unattainable because they are incompatible with certain natural or social laws or constraints. What can be done is to stretch the human life span, expand and deepen human knowledge, or — perish the thought — increase our power over other people. We can also develop some resources and introduce manual labor-saving devices, and multiply the opportunities for bounded and ephemeral happiness.

To ascertain whether an ideal is attainable we must start by building a conceptual model involving the laws and constraints of the system in question. If such a model suggests that the goal is within reach, we may design a plan capable of guiding the actions steering the system to the predetermined goal. In the case of plans designed to implement social ideals, i.e. goals to be reached by entire social systems, the former will involve not only laws and constraints, but also certain axiological or ideological tenets, such as those compressed in the catchwords "human rights" and "equal opportunity". Success and failure of the measures taken to implement the plan will be judged by the departure of the real state from the goal state. Such departures can often be estimated quantitatively. (For example, the Gini index equals the departure of the real distribution of wealth from equality: Recall Vol. 4, Figure 5.10.)

A feasible ideal must make room for imperfections and risks of several kinds: we must not equate "ideal" with "perfection" because in the real world there are no perfect systems, in particular perfect people. (Plato was quite right in restricting perfection to the realm of ideas.) In fact, even the simplest of all concrete systems, such as a deuteron or a hydrogen atom, is subject to distorting external perturbations. The number and magnitude of defects or imperfections (relative to some conceptual standard) increases with the complexity of the system. Crystallographers, solid state physicists, engineers, biologists, physicians, psychologists and many others are familiar with defects of various kinds: dislocations, impurities, deficits, excesses, deleterious mutations, organ dysfunctions, maladaptive behavior, etc.

Imperfection of some kind or other is unavoidable, but some imperfections can be removed. This holds in particular for our own moral imperfections. Even the somber Augustine admitted in his *De natura boni* (399?) that evil comes in degrees and is never complete for, after all, we are God's creatures, and God is the supreme good and therefore the source of no evil. Moreover he made some room for personal improvement though none for social reform. We, the children of the scientific and the industrial revolutions, as well as of a number of social revolutions and reforms, tend to believe in human perfectibility. However, after two world wars, on the brink of the third, and in view of uncounted economic, political and cultural crises, we cannot deny the boundless wickedness and stupidity of some of us. So, we seem to be increasingly prepared to admit

POSTULATE 2.5 (i) Every human individual and every human group is imperfect. (ii) Every human being and every social system is perfectible. (iii) Every human individual and every human group risks degradation.

It is up to us, as private persons and citizens, which is it to be: progress or decline. It all depends on the ideals we embrace: on whether or not these are feasible as well as rooted in genuine values. In particular, when conceiving of ideals, or of plans for their realization, we must remember that every desideratum has a dual (a counter-desideratum). Some desiderata are objectionable for not matching any basic needs or legitimate wants: this is the case with self-mortification, sadism, and domination. But in other cases a legitimate desideratum may be pursued in a single-minded way only at the risk of forfeiting other legitimate desiderata. For instance, the exclusive pursuit of health will turn a man into a health freak, that of love into a Don Juan, that of knowledge into a physical cripple, and so on: recall Sect. 2.1 on value conflict.

These considerations invite a postulate; but before we can formulate it we need

DEFINITION 2.7 An ideal is

(i) *feasible* (or *attainable*, or *realistic*) if and only if it can be reached with the available means or with means that can be made available within the foreseeable future;

(ii) *good* if and only if its realization contributes to the reasonable happiness of some persons without impairing the well-being of others;

(iii) *worthy* if and only if it is both feasible and good.

And now the assumption:

POSTULATE 2.6 (i) Every human being who wishes to succeed in attaining or retaining a state of reasonable happiness has some worthy ideals. (ii) Every progressive social group contains individuals who share some good ideals.

In other words, reasonable happiness (though not necessarily well-being) cannot be had without any good ideals. Likewise a society may be viable, though not progressive, even if none of its members has embraced any good ideals. We need good ideals only to improve our lot. Without them we are doomed to stagnation or decline. Cynics and pessimists, take note.

2.3 *The Summum Bonum*

Anthropologists, sociologists and historians teach that every society, and every social group within it, has its own predominant value system. The *summum bonum* (supreme good) of the armed forces is to maximize the number of casualties of their enemy (external or internal); that of the medical profession, to improve the health of everyone in the community; that of the business community, to make as much profit as possible without losing customers — and so on. How can then one talk of *the* supreme good of everyone in every society?

There is one value, though, that everyone holds across all social groups and societies: her own life and the lives of her loved ones. This attitude is not culture-bound but universal because it is inherent in our animal nature. And, since none of us is fully self-reliant, we normally care not only for our own lives and those close to us, but also for those of others, mainly those who can help us or whom we alone can help. If we are minimally rational we realize that what we ought to value most of all is the survival of humankind and, indeed, of the entire biosphere. No other goal, no matter how worthy or unworthy, can be attained unless there are human beings in a suitable environment. Therefore we propose

NORM 2.2 The survival of humankind ought to be the supreme good of all human beings.

Since in the present state of the world universal nuclear disarmament and environmental protection are required for human survival, they come next. See Figure 2.2. Everything else, even social justice and liberty, come thereafter. The ultimate option now is peace and a clean environment or extinction. The heroic slogans, such as "Freedom or death", and "Equality or death", have become obsolete in the face of the danger of extinction by way of nuclear war or further environmental

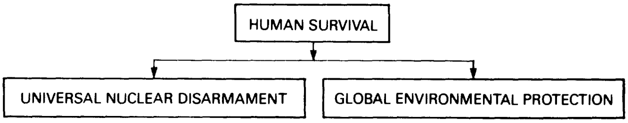

HUMAN SURVIVAL
UNIVERSAL NUCLEAR DISARMAMENT
GLOBAL ENVIRONMENTAL PROTECTION

Fig. 2.2. The *summum bonum*: the survival of humankind. The immediate necessities in order to secure the supreme good: universal nuclear disarmament and planetary environmental protection. The logical relation, in self-explanatory symbols, is $S \Rightarrow D \wedge E$.

degradation. However, this should not prevent us from hoping for more than survival. Once the survival of the species is assured through universal nuclear disarmament and global environmental protection we should have the will and resources to pursue other ideals, such as universal well-being and even reasonable happiness, as well as equity and freedom.

The adoption of Norm 2.2 requires a transvaluation of values — not the one Nietzsche advocated, though, because his own trinity, namely Individual life-Selfishness-Force, was far too close to the one that has brought us to the global crisis of our time, namely Pleasure-Profit-Power. We certainly need a revolution in our value systems if humankind is to survive. For one thing, we shall have to fasten some belts and loosen others. But such a revolution, far from burying all of the traditional values, should revive some of those lost in the frenzy of violence and consumption, such as solidarity, fairness, and kindness. However, this is a matter to be tackled in Chapters 4 to 6.

2.4 *Summary*

Whereas some values are mutually compatible, others are not. For instance, good health is compatible with education — nay, they favor one another. On the other hand speed and safety, quality and cost, expediency and democracy, and many other pairs, are mutually incompatible.

Value conflicts are unavoidable. When the two desiderata in conflict are legitimate, the conflict can be resolved by way of compromise. When at least one of them is not legitimate, the resolution calls for the elimination of one of the values. In this case we must alter, however slightly, our value system.

Any changes in values should be subordinated to our maximal ideals. These should be feasible and good. If feasible, they will make room for our own imperfections and those of the systems we handle, for nothing in the real world is perfect. As well, however good, all things human risk degradation. Lest we are willing to let things run down, we must protect our legitimate values by pursuing good ideals.

Whatever our ideals may be, we and our fellow human beings must stay alive if we are to realize them. Hence the supreme good is the survival of humankind. If this ideal is adopted as the *summum bonum*, then we must make alterations in our value systems. For one thing, our priorities cannot be profit and power but peace and environmental protection. Once these are secured we can argue about the rest.

CHAPTER 3

VALUE THEORY

In this chapter we shall discuss a number of problems concerning values that were either passed over or touched on only superficially in the previous chapters. More precisely, we shall be doing some analytic axiology or value theory rather than inquiring into the sources and functions of values, or declaring our own values. In short, in this chapter we shall analyze rather than evaluate.

Axiology may be defined by specifying the general concept of a research field (Vol. 6, Ch. 14, Sect. 1.2). We shall characterize it by the 10-tuple

$$\mathcal{A} = \langle C, S, D, G, F, B, P, K, A, M \rangle,$$

where, at any given moment,

C = the community of (professional or amateur) value theorists;

S = the society that hosts (and stimulates and inhibits) C;

D = domain = all the objects, real or imaginary, capable of being evaluated;

G = general background = the collection of ontological, epistemological, and moral principles taken for granted (assumed but not discussed) in a given axiological research project;

F = the collection of formal tools (usually only elementary logic at best) employed by members of C;

B = the collection of presuppositions (theological, biological, psychological, anthropological, sociological, etc.) employed by members of C;

P = the collection of (scientific and philosophical problems) concerning value in general and particular kinds of value;

K = the stock of axiological knowledge accumulated so far — not much at the time of writing;

A = aims = the elucidation and systematization of concepts and hypotheses about value;

M = methodics = conceptual analysis, hypothesis, argument, and the gathering of data.

Axiology has yet to find its proper place in the map of knowledge. As Ruyer (1952 p. 6) said, when picking a book on the subject one does not know whether he will find a treatise on theology, psychology, sociology, ethics, or metaphysics. This suggests that axiology, far from being a homogeneous and self-contained discipline, is a mongrel one characterized by multidisciplinarity. (See Vol. 6, Ch. 14, Sect. 3.2 for multidisciplines.) But the same fact also suggests that axiology may be regarded as branch of philosophy that, like any other branch, may borrow from other disciplines — in particular psychology and sociology, metaphysics and epistemology — and that, far from being self-serving, is cultivated mainly for the sake of other disciplines, in particular ethics and aesthetics, economics and political science. As a matter of fact we shall regard axiology as a branch of a science-oriented philosophy, rather than as a mixture of several disciplines all on the same level.

1. ANALYSIS

1.1 *Intrinsic and Instrumental*

Most value theorists teach that, whereas some things are valued for their own sake, others are valued for the sake of something else. The former are said to be *intrinsically* (or *inherently*), the latter *extrinsically* (or *instrumentally*) valuable. Thus health is said to be an intrinsic value, and medicine an instrumental value for helping us stay in or recover good health.

In our view there is no room for the intrinsic/extrinsic dichotomy because we regard all values as extrinsic and, more precisely, as relational properties. In fact, in our view whatever is valuable is so for some organism or group of organisms, in some respect and in some circumstance: change the individual or the group, the respect or the circumstance, and the value may change sign or even vanish altogether (Ch. 1, Sect. 1.1). So much so, that there would be no values in the absence of animals: whatever is valuable is so for meeting some need or want of some animal or group of animals (Ch. 1, Sect. 2.4).

What does make sense instead, and is very important indeed, is the *means-end* distinction — as long as it is not made into a dichotomy. If G is a goal, then M is a *means* for G if and only if M is necessary to reach G, i.e. $G \Rightarrow M$. If M is indeed instrumental to attain G, then whatever value is attributed to M may be said (elliptically) to be an instrumental value. But actually this is nothing but the value attributable to M.

Usually more than a single means or cause is required to bring about the desired effect of goal G. Calling $M = M_1 \wedge M_2 \wedge \ldots \wedge M_n$ the joint or successive occurrence of the causes M_1 to M_n necessary and sufficient to bring about the effect G, we express the means-end relation as a (contingent) equivalence: $M \Leftrightarrow G$. Caution: M and G are not occurrences but propositions representing such occurrences. More precisely, we have two systems: an agent (or her proxy), which shall remain anonymous, and her patient, a concrete system σ the state(s) of which the former is desirous and capable of altering. Call i the state of σ at the initial time t_i, and g the final state that the agent wishes σ to attain by time t_f on inducing a change m in σ. The propositions in question are

$M = \ulcorner m$ occurs in or to σ starting in state i at time $t_i.\urcorner$
$G = \ulcorner \sigma$ attains state g by time $t_f.\urcorner$
$M \Leftrightarrow G = \ulcorner \sigma$ reaches state g at time t_f if and only if m occurs in σ betwen times t_i and $t_f.\urcorner$

The means-end relation can then be analyzed in a systems-theoretic context. But it is not a dichotomy, for some states of affairs are both means and ends. For example, good health is both a goal and a means — e.g. to do one's work; friendship is desirable in itself (as a goal for the parties concerned) and as a means, e.g. to facilitate mutual help; on the other hand violence is bad both as a means (because it breeds more violence) and as a goal (because it jeopardizes the well-being of at least one person). In short, the notions of means and end are mutually complementary rather than mutually exclusive or opposing, for one and the same item may function as both means and end.

The short-sighted manager and the mainstream economist, as well as the unscrupulous politician and the Machiavellian political scientist, evaluate means and ends exclusively in terms of costs and benefits. For them, if $G \Leftrightarrow M$, then G is worth reaching by means of M only if the benefit accruing from G is greater than the cost of M. Such cost-benefit analysis is short-sighted because it omits the side effects that are bound to accompany any G, however good, and it overlooks the moral values of M and G. And this is morally wrong because either G or M or both may be wrong for violating rights or hindering the discharging of duties. It does not follow that cost-benefit analyses are wrong: far from it, they are indispensable in all spheres of individual and social life. Only, they are limited and must be supplemented with considerations of basic needs, legitimate wants, and moral constraints.

1.2 *Absolute and Relative*

In ontology one says that something is *absolute* if it exists by itself, not by virtue of or in relation to something else. In our ontology (Vols. 3 and 4) only the universe as a whole is self-existent, i.e. absolute: every part of it exists by virtue of its precursors. And, in addition to some intrinsic (absolute) properties, every thing in the world has some relational (e.g. frame-dependent) properties. Values are among such relational properties (Ch. 1, Sect. 1.1).

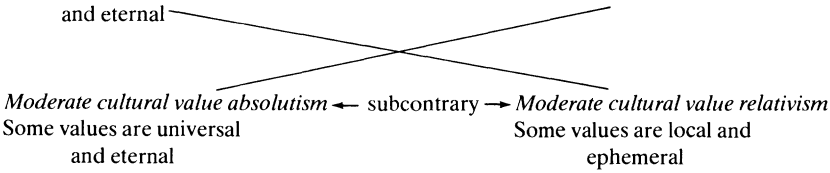
The axiological absolutists, from Plato and the Christian fathers to the phenomenologists Scheler, the early Hartmann, and Ortega, hold that all values are absolute (self-existent), and moreover, that they are timeless (above change) and universal (cross-cultural), for being ideal objects. Ontological value absolutists dispense with the evaluator, much as they dispense with the knowing subject — which in their view plays no role in value matters anyway. Consequently they handle axiology as a chapter of ontology, though usually without knowing it (Stern 1967 p. 121). Worse, with the exception of Plato they have no ontology to embed their axiology in. Even worse, by overlooking the evaluator — complete with his needs and feelings, knowledge and action — axiological absolutism leads inevitably to an authoritarian ethics. (See Stern 1936 and Frondizi 1963 for further criticisms.)

In his *Nichomachean Ethics* (Bk. 1, Sect. 5) Aristotle demolished ontological value absolutism, arguing that man does not know any goods in themselves, let alone *the* ultimate good which Plato rambled on about: he can only know good actions and good things. For example, the physician does not study Health but individuals in good, indifferent or bad states of health. In other words, the terms 'good' and 'valuable' are adjectives not nouns, for they denote properties — and in a naturalistic ontology properties do not exist by themselves but are possessed by concrete individuals. (Recall Vol. 3, Ch. 2, Sect. 5.2.) In sum, ontological value absolutism is untenable unless one admits Plato's idealistic ontology in its entirety — and foregoes the whole of modern science. A science-oriented value theory has got to be ontologically relativist. However, this does not entail that it must be anthropologically (or culturally) relativist as well. Let us explain.

We must distinguish *ontological* from *anthropological* (or *cultural*) value absolutism, i.e. the view that values are suprasocial and unhistorical, i.e. the same across all cultures and over the whole of history. This view, popular among the 18th century human nature theorists, has few followers nowadays — except of course among sociobiologists and social Darwinists. Most social scientists and not a few philosophers lean to anthropological value relativism, or even to situationalism, the view that our choice of values depends on the situation we are in. All of the historicists, from Vico and Montesquieu to Hegel and Marx, the functionalist anthropologists and sociologists, as well as the existentialists, deny the existence of impersonal, cross-cultural and unhistorical values. To them *all* values are particular, i.e. subjective or culture-bound, and likely to change in the course of time. Their devises are *Veritas et virtus filiae temporis*, and *Different situations, different values.* (For much detail see Brecht 1959.)

The cultural value absolutist fails to account for the actual variety in value systems found by social scientists around the world, as well as for the changes in values in the course of history. But the cultural value relativist ignores that all normal human beings value well-being, friendship, freedom of movement, activity, leisure, and much more — and that those who do not are regarded as abnormal. Nor is the cultural value relativist able to evaluate value systems, in particular to speak of progress or decadence in morals and customs. He must abstain from either praising or condemning any beliefs or practices prevailing in other societies or at other times, and he must be a conformist and perhaps even an opportunist in his own milieu.

Since cultural value absolutism and relativism are mutually contrary, not contradictory, there are alternatives to them. Indeed, it is possible to adopt an intermediate view combining elements of both, namely *moderate cultural value relativism*, according to which some values are universal and permanent whereas others are bound to place, time, and circumstance. (See Bidney 1953.) Clearly, *moderate* cultural value relativism is perfectly compatible with its subcontrary, *moderate* cultural value absolutism. The following square of oppositions summarizes the four (actually only three) views in question. The diagonals denote contradictions.



Cultural value absolutism ←—— contrary ——→ *Cultural value relativism*
All values are universal and eternal
All values are local and ephemeral

Moderate cultural value absolutism ←— subcontrary —→ *Moderate cultural value relativism*
Some values are universal and eternal
Some values are local and ephemeral

The absolutist may object that the moderate view is unsatisfactory because, although some values may be local and ephemeral as a matter of anthropological or historical record, they *ought* to be held by everyone under all circumstances. For example, well-being and peace ought to be desired universally at all times, even though the former is rejected by the flagellant and the latter despised by the conqueror. This objection is well taken but it does not prove absolutism anymore than the practices of slander and torture prove relativism. All it does is to remind us that axiology can be descriptive or normative — in which case it turns into morality.

In other words, the collection of all values may be partitioned in two different ways, according to whether they are held to be *de facto* or *de jure*. For example, reciprocity ought always to be practised, but as a matter of fact some of us sometimes fail to return services done us. Thus reciprocity is absolute (or rather universal and perpetual) *de jure* though not *de facto*. On the other hand, *pace* Kant, truthfulness is only a situational value, for not everybody is or ought to be truthful in all circumstances — e.g. when interrogated by the secret police. Thus in addition to the two previous dichotomies — ontological and anthropological value absolutism and relativism — we have the following tetrachotomy in the set of all values.

Universal de jure (e.g. reciprocity)	*Universal de facto* (e.g. self-preservation)
Local de jure (e.g. all life for Buddhists)	*Local de facto* (e.g. politeness among the Japanese)

Only the universal *de jure* values, such as well-being, solidarity, fairness, and peace, come close to deserve being called 'absolute' — provided they are not regarded as self-existent. But even they are historical rather than unhistorical, as shown by the hard times that solidarity, peace, and other values, are having.

The preceding distinctions should not be underrated as mere academic hair-splitting. Indeed, it is arguable that humankind has gotten itself into trouble because far too many people have adopted an ethnocentric stand, mistaking their own local *de facto* values for universal *de jure* ones, while at the same time giving certain wants priority over basic needs. Because of that value intolerance we are involved in a number of dangerous conflicts, and because of the widespread wish to gratify wishes regardless of needs, resources are being wasted on a huge scale. But the matter of subjective values deserves a separate section.

1.3 *Objective and Subjective*

Ontological value absolutists are axiological objectivists, i.e. they believe that values are subject-independent. On the other hand most ontological value relativists are axiological subjectivists, i.e. they believe that values "are not part of the fabric of the world" (Mackie 1977 p.15). Protagoras, Nietzsche, Weber, Russell, Perry, Sartre, Mackie and many others have defended axiological subjectivism. Who is right? The answer to this question depends critically on the definition of the concepts of objectivity and subjectivity with reference to values.

The epistemological concept of objectivity is clear: Something is said to be *objective* if it exists independently of any knowing subjects, i.e. whether or not it is known to someone. Thus the proverbial thunderclap in a desert is objective whether or not anyone happens to hear it; moreover it can be recorded by an instrument. We cannot use the same definition in value theory because there are no goods except for evaluating subjects: values emerged and evolved along with the animal kingdom long before animals capable of knowing, i.e. knowing subjects, appeared.

All values are subject-rooted even though not all of them are subjective. Likewise motility and vision occur only in animals yet they are perfectly objective in the epistemological sense of 'objective'. On the other hand beauty and charm are subjective in the sense that their attribution is a matter of individual appreciation. We generalize and stipulate, in line with Ch. 1, Sect. 3.1, that any values rooted to basic needs or legitimate wants are *axiologically objective*, and the rest subjective. Thus, nutrition and solidarity are axiologically objective, whereas aesthetic pleasure and ridicule are subjective. The admission that some values are axiologically objective whereas others are subjective is clearly at variance with both axiological objectivism and subjectivism.

Let us stress the difference between epistemological and axiological objectivity. In epistemology, whatever is subjective is relative (to the knowing subject); equivalently, whatever is absolute is objective. The converse is false, for there are relative items, such as distances, durations, and energies, that are both objective and relative (to the reference frame and a fortiori to the knowing subject). On the other hand in our axiology all values are relative to the evaluator, but while some are objective others are not. Equivalently: There are objective standards or canons for determining whether a given item is really good or bad for someone's well-being. On the other hand there are no standards or canons in the case of likes and dislikes that fail to originate in basic needs or legitimate wants.

To state that some value judgments are objective and others subjective amounts to holding that, whereas some value judgments are objective, others are not. Thus "Milk is good for all newborn mammals" is an objective value judgment: it is true. On the other hand "I like milk" is subjective — yet it happens to be true of the present writer. Moral: In matters of valuation subjectivity is not a falsity indicator. Nor is intersubjectivity or unanimity a truth indicator: Sometimes we all make the same mistake, e.g. that of believing in the educational value of strapping or in the moral virtue of capital punishment.

One problem with subjective values, which the objective ones do not pose, is that of authenticity. We cannot fake primary or secondary values because the objects possessing them meet basic needs. On the other hand we may fake tertiary or quaternary values — particularly the latter — because they stem from desires or even fancies. Thus people who hanker after junk food, instant music, or tachiste ("action") painting, attribute them pseudovalues. We shall stipulate that an object is attributed a *pseudovalue* if it is attributed a value that it lacks.

To be sure, beauty is subjective, but genuine art involves certain canons that have to be learned, and an aesthetic sensibility that must be cultivated. It is typical of pseudoart that it is produced (*a*) by individuals who have not bothered to undergo a rigorous artistic apprenticeship, (*b*) for profit — of manufacturers and their managers and *marchands* — rather than for fun or for the sake of an artistic ideal, and (*c*) for mass consumption, particularly by the uneducated. Bogus art is often more dependent on technology than on inspiration; its appreciation depends more on commercial promotion than on education; it is short-lived; and it has a low, often negative, cultural value despite its high commercial value. It is a triumph of audacity over responsibility, of improvisation over tradition, of business over culture. But unlike junk food, which is a poor substitute for wholesome food, junk art is a toxic travesty that contributes powerfully to the decline of genuine culture. Addiction to bogus art renders people insensitive if not allergic to the authentic article. And, because much of junk art expresses violence or selfishness, it contributes to moral and social decadence. However, it would be mistaken to censor it: a state that adopts a particular aesthetic view suffocates all artistic creativity. The only effective antidote is education.

The preceding distinction between authentic and bogus values makes little psychological sense. Indeed, we tend to value positively whatever gives us pleasure, even if it is objectively worthless or even disvaluable; likewise we tend to reject as disvaluable whatever gives us pain, even if it promotes our health or social standing. It takes some knowledge and willpower to discard pleasurable pseudovalues and admit painful valuables. Because of such discrepancies between what we like and what is objectively good, it would be mistaken to take the psychology of reward and punishment as the basis for value theory and action theory (as Homans 1967 tried to do).

Plotinus asked whether we desire what we value, or value what we desire. Plato, Aristotle and others held that we desire or commend things, actions, and other items, for being good. On the other hand the emotivists believe that we call 'good' whatever we regard as desirable or commendable. Each of these contending views holds part of the truth. The truth is that Plotinus's option is not an alternative, for it is not an instance of the excluded middle law. Hence we may admit one of the theses in certain circumstances and the other in others. Sometimes we desire X because we know or imagine that X is good; but we may have to know or imagine that X is good because X attracts us. Evaluation may then precede or follow a conative state, and desire may precede or follow evaluation. We are then confronted with a process wherein objective and subjective values intertwine rather than forming a circle.

1.4 *Individual and Social*

Values may be classed into individual (or personal) and social (or collective). The former can only be realized by individuals, the latter by collective action. Thus amiability and uprightness are individual values, whereas peace and social justice are social values. The social values may in turn be divided into local and universal (Sect. 1.2). The former depend on idiosyncrasies of a community or region, whereas the latter are independent of such peculiarities. (See Stern 1967.) For example, whereas environmental protection and international cooperation are universal values for benefiting everyone, industrialization is only valuable for underdeveloped regions: the developed ones are in some respects overindustrialized.

Individualists admit only personal values, which they tend to regard as both subjective and permanent, and they deny the existence of social values. Thus Buchanan (1954), in his important criticism of Arrow's seminal work on social choice (1951), stated that "We may adopt the philosophical bases of individualism in which the individual is the only entity possessing ends or values", and in which no social entity is presumed. (See also Buchanan 1975a.) While it is true that only individuals can evaluate, it is equally true that individuals can and in fact do evaluate anything, particularly societies and their subsystems. Hence Arrow is not wrong in speaking of a "social value scale" as different from a personal one. Where he does go wrong is in assuming that such a social ranking can be derived from individual preferences: whether or not a given state of society is better than another is for social scientists to decide, not for an opinion poll to find out.

Individualists are so obsessed with individual values and rights, such as liberty and private property, that they fail to acknowledge social values, such as peace, environmental protection, popular participation, and public education. The legal consequence of this outlook is the thesis that the only justifiable cause for limiting individual liberty (e.g. by imprisonment) should be the prevention of harm or serious offense to persons other than the actor. (See e.g. Mill 1859, Feinberg 1984, 1985.) If interpreted literally, this view is indifferent to the wanton destruction of public property, warmongering, or the use of relief funds for political blackmail. And of course it favors shrinking the role of the state to that of a nightwatchman.

The sociological and axiological holists (e.g. Vierkandt, Durkheim, Bouglé, Lévy-Bruhl, Westermarck, Benedict, and Herskovists) hold that society is the source and support of *all* values. Hence they tend to deny the existence of individual (biological and psychological) values as well as of universal ones. This denial is mistaken for at least two reasons. One is that individual values are prior to all others in the sense that we cannot form or hold any social values unless we are alive and well. A second reason is that, if individual values are denied or underrated, then persons are treated as mere tools for social groups, e.g. authoritarian governments. This is not only morally wrong, as Kant taught us, but is also socially destructive, because people do not function efficiently unless they meet their basic needs and legitimate aspirations.

We have criticized sociological individualism and holism elsewhere (Vol. 4, Ch. 1, Sect. 4.2, and Vol. 7, Ch. 4, Sect. 1.1). This criticism carries over to axiological individualism and holism. We propose *axiological systemism* instead. According to this alternative (*a*) there are personal as well as social values; (*b*) whereas all social values are objective, some personal values are subjective; (*c*) personal values are constrained by social values and conversely, and this mutual constriction is a frequent cause of value conflict; (*d*) in a viable society social values are instrumental in facilitating the realization of personal values; (*e*) in a progressive society value systems are subject to criticism and change.

1.5 *Fact and Value*

Vulgar materialists (physicalists) make no room for values in their world view, even though they cannot avoid making value judgments in their everyday life. On the other hand idealists, in spite of being part of the only world there is, detach values from facts and moreover they place the former above the latter. They do not tire of warning us not to mistake value judgments for empirical ones and they reject out of hand any attempt to justify valuation with the help of knowledge — "one of the strangest aberrations ever to visit the mind of man" (Lewis 1946 p. 366). In short, for them there is an unbridgeable gap between fact and value, hence between science and axiology.

Non-idealists refuse to reify values. To them there are no values in themselves but only objects that certain organisms value for themselves or their groups. (Recall Ch. 1, Sect. 1.1.) Accordingly they regard value as an abstraction from valuation, and the latter as a special kind of fact. In particular, materialists view conscious valuation as a brain process partially conditioned by social circumstances as well as by inner biological and psychological urges. In short, for naturalism there is no fact/value barrier, and values are accessible to science. (See e.g. Skinner 1971, Ch. 5.) This does not entail that we can define *every* value in biological or psychological terms. In particular, it would be mistaken to adopt Skinner's identification of "good" with "positive reinforcer" and "bad" with "negative reinforcer", for this would leave out all social values and would lead to a hedonistic morality.

Values, then, have a comfortable place in any non-idealistic and non-physicalistic ontology. In such an ontology values, or rather valuations, are not autonomous ideas but processes occurring in animals. In this perspective a value judgment, i.e. a proposition of the form "*V* is valuable", is construed as being short for "There is at least one animal for which objects with property *V* meet some of its needs or wants". (Recall Ch. 1, Sect. 2.4.)

Ours is then a naturalistic axiology that dovetails with our naturalistic ontology. We indulge deliberately and cheerfully in what idealists and intuitionists call the "naturalistic fallacy", for we naturalize some values and socialize others. In this regard our axiology is in the tradition of such diverse thinkers as the Greek atomists and the Stoics, Aristotle, Spinoza, Hobbes, Hume, Bentham, Marx, Durkheim, Bergson, Russell, and Skinner. But, whereas Hume, Bentham and Skinner admitted only psychovalues, and Hobbes, Marx and Durkheim recognized only sociovalues, we recognize the biological source of values as well.

Moreover, far from asserting that values are just epiphenomena, we admit their active role as guides to action, particularly action of the deliberate kind. (Consciousness, with its dual role of dashboard and steering wheel, is parallel.) The main flux goes from needs or wants to knowledge and valuation, thence to action. In fact, rational beings guide their behavior by (true or false) knowledge as well as by (right or wrong) valuation. But they take cognizance of the outcome of their action and, unless they are hopeless dogmatists, such knowledge is bound to force some corrections in both their fund of knowledge and their value system, which in turn is likely to put constraints or wings on their desires. See Figure 3.1.

In short, there is no fact-value gap. Value systems have concrete (biological, psychological or social) sources, they exist as biological dispositions or as brain processes, and they end up by being "embodied"

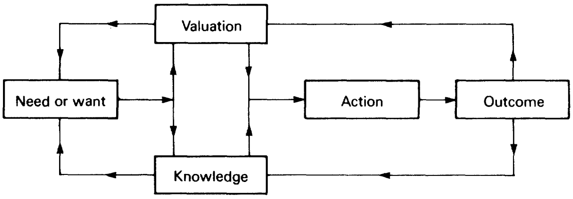


Fig. 3.1 Values as links between needs and wants, on the one hand, and actions on the other. Note the feedback loops from the outcome of action to knowledge and valuation, thence to the latter's source.

(pardon the Platonism) in actions and their outcomes. The fact-value gap is only a figment of the imagination of unworldly philosophers.

1.6 *Summary*

Value theory has traditionally been dominated by the absolute/relative dichotomy, where in turn "absolute" has been identified with "self-existent", "intrinsic", or "objective", whereas "relative" has been identified with "other-related", "instrumental", or "subjective". We have attempted to elucidate these concepts and to argue for a position midway between axiological absolutism and relativism.

We have denied that there are any absolute and intrinsic values, and have held that the intrinsic-instrumental distinction should be translated into the ends-means one. However, this distinction does not amount to a dichotomy, for one and the same item, such as well-being or freedom, may be both a means and an end. The objective-subjective distinction stays with the qualification that some desires, though subjective experiences, are objective values rather than mere whims. Examples: the desires to be useful and to improve one's knowledge. Likewise we must distinguish personal from social values, while admitting that the realization of each of them constrains that of some values in the complementary category. Finally, we have rejected the fact-value dichotomy, holding instead that evaluations are facts. Whether value judgments can be true shall be looked into forthwith.

2. VALUE AND KNOWLEDGE

2.1 *Semantic and Epistemological Status*

According to popular wisdom all value judgments are entirely matters of taste or opinion: unlike factual or empirical judgments, they are not part of knowledge and they have no truth value. The emotivist doctrine of values and morals, which has been part and parcel of positivism since Hume, is little more than an explicit formulation of that popular view. (See e.g. Ayer 1936, Stevenson 1944.)

If values and norms have only an "emotive meaning", then they are cognitively meaningless. This follows on the verifiability theory of meaning, according to which the meaning of a linguistic expression is the way it is put to the test. If all value judgments are nothing but a matter of emotion or of taste, then they are untestable and therefore meaningless. But, even assuming that all value judgments were untestable — an assumption to be questioned below — it would not follow that they are meaningless, and this simply because the verifiability theory of meaning is false. Indeed, before imagining a test capable of finding out the truth value of a proposition, we must grasp at least part of the meaning of the latter (Vol. 2, Ch. 7, Sect. 51.). Therefore the view that value judgments are meaningless is false. The founder of the Vienna Circle admitted this point (Schlick 1930), and his follower Kraft (1937) wrote a book on values. So, the view in question has become a mere footnote to the history of philosophical extravagance.

Evaluation may involve emotion — but so does cognition. Evaluation is a kind of cognition (Lewis 1946). A rat that presses a lever for more food pellets of a certain kind knows that the food is good for him — but he will revise this evaluation if the last time he sampled food laced with poison. A person who has learned that keeping promises is necessary to stay in good social standing knows something about certain values, and she has come to know it on the strength of experience and reason. And a person who realizes that she has made mistakes in value attributions will correct the latter much in the same way as she corrects perceptual or calculation errors.

Value judgments can be justified or criticized, rather than accepted or rejected dogmatically, when they are rooted to basic needs or legitimate wants. In this case they can be shown to be true or false. Thus consider the following propositions.

(i) Freedom is good for allowing us to exercise our rights.
(ii) Honesty is good for promoting cooperation.
(iii) The partaking of meals is good for favoring social cohesion.

What is arbitrary about these value judgments? They can be justified or criticized in the context of social science. Even if someone were to disagree with any of the above propositions, he would do so because he regards it as false.

Of course some value judgments *are* subjective. For example, it is largely a matter of taste whether a given work of art *is* good. (However, recall what was said in Sect. 1.3 about pseudovalues.) But in this case all we have to do, in order to avoid fruitless controversy, is to abstain from claiming that a certain piece is objectively good: we can retreat to the more modest 'I like it', which is, in principle, testable. For example, a psychologist employing physiological techniques will have no difficulty in finding that it is true that I dislike the music of Bartok, Britten, and the like. But this is not an aesthetic judgment: it is merely a psychological truth about myself in relation to those composers.

If all value judgments were devoid of cognitive content it would be hard to learn them and even harder to correct them in the light of reason informed by experience. Yet we keep upgrading (or downgrading or merely shifting) our value systems — e.g. as we age we may become either more sophisticated or more sensitive to basics. Unlike the emotivist and the intuitionist philosopher, ordinary rational beings demand that reasons for or against non-aesthetic value judgments be given. Only utter conformists passively receive whatever values are fashionable.

A positive rationalist may demand, in the Aristotelian tradition, that a "practical reason" be given for or against every value judgment and every norm. (See e.g. Korsgaard 1986.) Regrettably the notion of practical reason is unclear to the point of coming close to being self-contradictory. Only causes, motives, desires, and the like can be practial: reasons are conceptual. Therefore instead of saying that someone does (or ought to do) something for some "practical reason", we had better say that she does it for some cause, motive, or desire.

As for the negative rationalist, taking his cue from Popper he may demand that every value claim be criticizable (e.g. Albert 1985). This is correct but not enough: as in the case of hypotheses we need something positive and empirically accessible to justify our valuations. Basic needs and legitimate wants supply just such positive grounds. Moreover, since such deficits can be the subject of scientific investigation, the corresponding values can be discovered rather than invented. This view, orthogonal to subjectivism (in particular emotivism and conventionalism), squares with our realistic and scientistic epistemology (Vols. 5 and 6). Some more on this anon.

2.2 *Values and Science*

The received wisdom concerning the relation of axiology to science is that they are disjoint because (*a*) value claims cannot be justified on the basis of empirical data, and (*b*) science ignores values. We criticized opinion (*a*) in the preceding section, so we may now concentrate on opinion (*b*).

In discussing the relation between science and values we must distinguish the values inherent in scientific research (e.g. honesty) from those attributed by scientists to some of the things they study (e.g.

selective value). The existence of the former is obvious: scientists evaluate their own work all the time. They evaluate their own research projects, problems, hypotheses, theories, data, methods, and those of their peers and students, as being good, mediocre, or bad. For example, experimental designs are rated as good or bad depending on whether they involve adequate controls, precise hypotheses, and precise techniques — or not. Theories are rated as good or bad according to their explanatory and predictive power as well as to the way they harmonize with other theories. Entire approaches are evaluated as a matter of course. For example, contemporary historians believe that their predecessors underrated the environmental, demographic and economic factors, and overrated the importance of political, diplomatic and military events. In short, science has its *endovalues.*

But of course science can also be evaluated from the outside. Thus methodologists evaluate entire fields of knowledge as being good science, bad science, or neither. Technologists evaluate scientific methods and results as useful or useless for their own work. Statesmen evaluate economic and political theories as relevant or irrelevant to statecraft. Sociologists and historians weigh the importance of science in different cultures. Some political activists blame science for the sins of technologists and their employers — and so on.

In addition to the value judgments about science, made by scientists and nonscientists, we have the attributions of value to objects of scientific research, or *exovalues.* To be sure, physicists and chemists make no such judgments unless they work as technologists. Atoms, fields, stars, chemical reactions, and the like, are neither good nor bad: they are value-free. But the situation alters radically the moment we pass from physics and chemistry to biology or to the social sciences. Indeed, the biologist may say that certain genes are defective, that proteins are good for all organisms, and that most viri are bad for them; that genetic variety is good for species survival, and that species variety is good for the biosphere; that industrial pollution is bad for nearly all organisms, and so on. (Recall Ch. 1, Sect. 2.1 on biovalue.)

But of course the science, or rather semiscience, where value concepts crop up most often is economics, the parent of modern axiology. To begin with, economics deals with goods, i.e. items supposed to be useful to some people. Then, it studies their values in use and their values in exchange (or prices). The classical economists, in particular

Smith and Ricardo, as well as Marx, attempted to build a general theory of economic value, namely the so-called labor theory of value. According to this theory, or rather hypothesis, the real measure of the value in exchange of a commodity equals "the toil and trouble of acquiring it" (Smith 1776 Bk. I, Ch. V). Because there is no objective measure of "toil and trouble" — as Smith himself admitted — the hypothesis has remained untested. For the same reason some modern economists, such as Robinson (1962), have denounced the concept of value as non-scientific. Yet, in one guise or another value concepts keep cropping up in economics, which would be pointless if commodities were as value-free as galaxies.

If Max Weber had paid more attention to economics, or had known that biologists make value judgments all the time, he might not have waged his famous campaign for the value neutrality of the social sciences. True, the social scientist is supposed to make objective and if possible true statements, but some of these may be value judgments. Examples: "Peace is better than war", "The arms race is bad for the economy, the polity and the culture", "Birth control promotes the quality of life", "Land reform is good for everyone except landowners", "The right to work is more important than the right to vote", "Free trade is good for the industrialized nations, bad for the less developed ones", "Unemployment and inflation are bad for everyone", and "Power corrupts".

Whether or not any of the preceding statements is true, is immaterial at this point. The point is that they are value judgments made by social scientists. Moreover they are not mere expressions of likes or dislikes: they can be put to the empirical test. For example, social and economic statistics show without doubt that the arms race is bad for society because it decreases the social expenditures (particularly in public health and education), diverts funds from civilian business, causes unemployment, slows down technological progress (except for the technology of mass murder), threatens liberty and democracy, and augments the risk of international conflict. (See e.g. Sivard 1987, Bellon & Niosi 1988, Brown Ed. 1989.)

In short, the social scientist cannot avoid making value judgments. Intellectual honesty demands that he declares openly his own values (Myrdal 1942, 1969). It also demands that he makes an effort to justify them instead of stating them dogmatically: this, the concern for empirical tests, is precisely one of the differences between science and nonscience with regard to value statements. (For a history of the relation between science and values see Graham 1981.)

2.3 *Values and Technology*

The task of the technologist is to design artifacts — things or processes, inanimate or living — or to supervise their construction or maintenance. (Recall Vol. 7, Ch. 5.) Since artifacts are supposed to be of value to somebody, the technologist must evaluate not only his own work in point of technical quality: he must also evaluate the possible social function of the artifact he designs. Furthermore he must evaluate the natural and social resources required to materialize his designs.

Technological value judgments are supposed to be objective. Thus when an engineer states that device A is *better* than device B of the same kind, he means that A is more efficient, reliable, lasting, or useful than B. And he is supposed to *back up* this value judgment with calculations and tests. Likewise when a manager, applied economist, applied political scientist, or civil servant designs a social organization — such as a firm or a public service —, an economic policy, a social program, or a political campaign, he is supposed to do so on the basis of an objective study, as well as on an estimate of costs and benefits. Every technological design is supposed to be evaluated at least twice: before and after its implementation. Before, to ascertain whether it promises benefits or some kind; after, to find out whether it works as promised, or whether it needs revision or even scrapping.

The technologist is apt to evaluate anything within his reach. Not even physical and chemical systems will escape his evaluating gaze. For example, whereas to the physicist heat is just a kind of energy, to the engineer it is a degraded form of energy because it cannot be wholly reconverted into mechanical or electromagnetic energy. Therefore to him an increase in entropy is a decrease in the value of energy; if socially conscious he will preach minimizing the increase in entropy. And the engineer, as well as the environmentalist, think of the release of heat into the environment as thermal pollution — and rightly so. Where the scientist sees a chance event, the technologist may see a risky one. Where the former may limit herself to estimating the probability of the event in question, the technologist may attempt, in addition, to measure the risk it presents. (A possible measure of the risk incurred when using means M to attain goal G is the product of the improbability of attaining G by the cost of M, including the cost of the negative side effects of both M and G.)

Few items highlight the difference in the valuational attitudes of the scientist and the technologist as garbage. To the basic physicist or chemist the concept of rubbish makes no sense, whereas it does to the biologist, social scientist, technologist, and others. Waste disposal has become a major technological and social problem, what with every American producing about ten tons per year, and the US industry about 250 million tons of hazardons waste per year. A first problem is that of defining the very concept of garbage, which in ordinary life we associate with dirt, ugliness or even baseness. The engineer tends to think of rubbish as a misplaced thing, for what is waste here might be raw material or energy somewhere else. For example, the heat and even part of the smoke released by a smoke stack might be used in greenhouses, and most of the toxic wastes can be broken down into useful chemicals, either by chemical or by physical methods. Nothing, not even sewage, is inherently dirty or wholly useless; even radioactive waste can be used, and is actually being used on a small scale, to kill bacteria and parasites in fruits and vegetables. Hence the rational way of "disposing" of garbage is to recycle it, not to dump, bury, or burn it. Technologists are more effective transvaluators of values than value theorists. Still, neither is likely to come up with a technique for the disposal of cultural rubbish. It would seem that the only effective solution to the problem of the mass production and consumption of this kind of rubbish is popular (not just elitist) education on all levels.

2.4 *Summary*

Some value judgments are mere expressions of taste, caprice, or passion: they lack objectivity. But others are objectively true or false for being justifiable or criticizable in the light of ordinary experience or of scientific findings. This holds in particular for the value judgments of the form "If X meets a need of Y, then X is valuable for Y even if Y does not desire X".

Science is not alien to values. First, because the scientific investigator holds certain psychological values such as curiosity, ingenuity and honesty; semantic values such as clear meaning and maximal truth; and methodological values such as testability and precision. Second, because developmental and social psychologists, as well as anthropologists, sociologists, economists, and political scientists, are supposed to study value systems in a scientific manner. Third, because social scientists, particularly those of the applied kind, cannot help making value judgments, such as "Poverty is degrading", "Greed breeds violence", and "Violence is socially disruptive", which happen to be true.

Technology is even more value oriented than science. In the technologist's purview everything, be it design, resource, or final product, has a value tag. Technological designs are supposed to be not only technically correct but useful to somebody and, ideally, harmful to nobody: they are supposed to be instrumentally good, i.e. to meet the needs or wants of some people. The facts that contemporary technology is largely geared to mass murder and to creating wants rather than to satisfying basic needs are unfortunate but not to the point. The point is that technology has become the most powerful conceptual tool for the realization of values both positive and negative.

In short, value matters are matters of fact and knowledge, not just of taste or emotion. Valuation is a mode of cognition — but of cognition of ourselves rather than of the external world. A value judgment of the form "X is valuable to Y" says nothing about the object X of valuation but it does make a (true or false) statement about the subject Y of valuation in relation to the value-bearer X.

3. VALUE MEASURES AND CALCULI

3.1 *Objective Value of a Generic Item*

Some values, such as justice, friendship, and beauty, are non-numerical, but they can often be ranked. Others, such as nutritive value, standard of living, and participation, can be quantitated. In the former case one speaks of "ordinal scales", in the latter of "cardinal (or metric) scales". In still other cases one may suspect that the values in question could be quantitated if only they were subjected to suitable theories and if adequate indicators were found. For example, until a few decades ago the performance of an artifact was evaluated intuitively, but nowadays optimization theory defines several performance indices, such as the so-called integral square error index. These are functions of the differences between the actual and the desired values of a certain property of an artifact. Whatever decreases such error is deemed to be good, and whatever minimizes it is said to be best.

We shall not incur the utilitarian error of presupposing uncritically that everything can be evaluated numerically. Nevertheless we shall assume that *basic* values, or rather the items that meet basic needs, can be assigned numbers. In line with Definition 1.14, we shall assume that the (basic) value of an item is the degree to which it meets a basic need. And we shall add that the (basic) disvalue of an item is the degree to which it generates a basic need. In either case an item may be a thing, a state of a thing, a process, an idea, or what have you. Thus, if being able to occupy a dwelling is assigned value 1, being evicted from it without having the possibility of moving to another may be assigned value -1, and being allowed to occupy only half of it may be attributed value $\frac{1}{2}$. See Figure 3.2.

More precisely, we propose

POSTULATE 3.1 Let x be a kind of items capable of satisfying or generating a basic need y of a given animal in a given state, $A(x)$ the amount of x available to the said animal, and $N(x, y)$ the amount of x that the animal needs to fully satisfy y. The value of x for the given animal in the given state relative to y is

$$V(x, y) = \operatorname{sgn}(x, y) \cdot \left[1 - \frac{A(x)}{N(x, y)} \right],$$

where

$$\operatorname{sgn}(x, y) = \begin{cases} 1 \text{ iff } x \text{ satisfies } y \\ -1 \text{ iff } x \text{ generates } y. \end{cases}$$

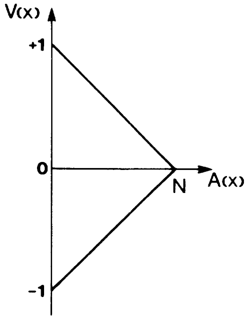


Fig. 3.2. Value is proportional to deficit or deprivation: See Postulate 3.1.

The preceding assumption allows us to define "good" and "bad" with respect to basic needs:

DEFINITION 3.1 If $V(x, y)$ is the value of item x relative to basic need y, then

(i) x is *good* relative to $y =_{df} V(x, y) > 0$;
(ii) x is *bad* relative to $y =_{df} V(x, y) < 0$;
(iii) x is *indifferent* relative to $y =_{df} V(x, y) = 0$.

The comparative concepts of better and worse are definable in terms of the quantitative concept:

DEFINITION 3.2 If $V(x, y)$ and $V(x', y)$ are the values of items x and x' respectively relative to basic need y, then

(i) x is *better than* x' relative to $y =_{df} V(x, y) > V(x', y)$;
(ii) x is *worse than* x' relative to $y =_{df} V(x, y) < V(x', y)$;
(iii) x and x' are *equivalent* relative to $y =_{df} V(x, y) = V(x', y)$.

Clearly, the better and worse relations are irreflexive, antisymmetric, and transitive: i.e. they are strict partial ordering relations.

Finally, we define "preference" in terms of "value", namely thus:

DEFINITION 3.3 An animal b given free choice between items x and y *prefers* x to $y =_{df} b$ values x more than it does y.

Actual choice, when deliberate, presupposes preference. (Random choice is no exception, for it is no choice proper: when we let the coin or the die decide for us we surrender our choice ability.) Yet to an operationist choice is identical with preference: he mistakes the indicator (choice) with the indicated (preference). This is not only a conceptual muddle: it is also empirically misleading, for indicators, particularly in psychology and social science, are often ambiguous. Thus an animal under restraint or compulsion is unlikely to choose what it prefers: if left no choice but to perform a certain action then it has no proper choice at all. Sheer ignorance can have the same effect as compulsion. Thus Buridan's ass died of hunger for ignoring that, if two haystacks are equivalent, then it does not matter which one is chosen.

The conflation of "good" with "preferred" is frequent in all of the so-called policy and decision sciences. For example, the Pareto principle

P1 If A benefits some people without harming anyone, then A is good.

is sometimes stated in the form

P2 If everybody prefers A to B, then A is better than B.

But clearly the two statements are inequivalent: Whereas *P1* is acceptable, *P2* is not, for whereas benefits can be checked objectively, some preferences are subjective or even irrational. Thus the popular preference in North America for "strong leadership" over participatory democracy is no proof that the former is better than the latter.

We shall not subject the preference relation to the standard von Neumann and Morgenstern (1947) axioms, because these have been found to be not only empirically false but also doubtful guides to rational action, hence with no certain normative value. (See e.g. Allais & Hagen Eds. 1979, Anand 1987.) We submit that only psychological research, not a priori speculation, can determine, say, under which conditions an animal's preferences are actually transitive, hence "rational". The mathematician and the philosopher lack the empirical tools required to determine the psychological laws of preference. (See May 1954, Tversky 1969.)

Note finally that the preceding definitions, as well as Postulate 3.1, hold only for basic needs; they are not true of all desiderata, whether or not these are legitimate. Notice also that, to the extent to which the above analysis is correct, it refutes the intuitionist theses that (*a*) the value concept is simple and (*b*) the good is always knowable preanalytically and a priori. To be sure, in some cases it is easy to tell good from bad: instinct or experience may suffice. But such cases pose no interesting axiological problems. We need axiology when intuition fails us.

3.2 *Objective Value of Systems and Processes*

So far we have dealt with generic items — things or processes, natural or artificial — capable of meeting basic needs. The value concept can be further analyzed in systems-theoretic terms; in particular, we can attribute values to complex things and to processes occurring in them. Such analysis will reinforce the links between our axiology and our ontology (Vols. 3 and 4).

We start off with the intuitive idea that the more valuable systems are the more complex or advanced ones (L. Puelles, personal communication). Thus, we say that a vertebrate is more valuable than any invertebrate, and a combine more valuable than any shovel. However, there are counterexamples to this intuitive idea: for instance, a complex aggressive army is worse than a police force. To begin with, for a system to be valuable it must meet someone's basic needs or legitimate wants, on top of which it must be viable or robust rather than maladapted and readily disintegrable. Therefore we assume

POSTULATE 3.2 The more valuable of two good systems is the one that is the more complex and robust.

Example 1 An expecting mother, if likely to survive childbirth, is more valuable than her fetus.

Example 2 A good society is better than any of its components. (On the other hand some individuals of a bad society may be better than the latter.)

We now turn to processes, in particular to those occurring in the supersystem composed of the evaluator and the object of evaluation. For the sake of simplicity we shall restrict our analysis to the case where the supersystem can change from one state to another through the intermediary of a single state — which, however, may be different in different circumstances. In other words, we shall lump the transient states into a single intermediate state or means m lying between the initial state i and the final (or goal) state f of the supersystem in question. In short, we shall consider a process $p = \langle i, m, f\rangle$ and shall assume that, at least in some cases, it and its constituent states can be evaluated quantitatively. A simply assumption for the value of such process is

POSTULATE 3.3 Consider a supersystem composed of an evaluator and the things over which he has control. Call $p = \langle i, m, f\rangle$ the process whereby the system goes from its initial state i to the final state f passing through the intermediate state m, and assume that every one of these three states has been evaluated according to Postulate 3.1. The value of $p = \langle i, m, f\rangle$ is

$$V(i, m, f) = \tfrac{1}{2}[1 + V(m)] \cdot [V(f) - V(i)].$$

In words: The value of a process under (direct or indirect) animal (e.g. human) control is proportional to the difference between the values of the final and the initial states, the proportionality constant being a linear function of the value of the means employed to effect the transition. This factor does not occur in utilitarianism or in utility theory, for which only the difference between the values of the goal and the starting point matters. In our view the means are just as important as the goal, to the point where the use of the wrong means may bring the entire effort to nought. The following examples will bring this point out.

1 *Success: Increase in value with valuable means*

$$V(i) = 0,\ V(m) = \tfrac{1}{2},\ V(f) = 1,\ \therefore\ V(i, m, f) = \tfrac{3}{4}.$$

For utilitarianism, $V(i, m, f) = 1 - 0 = 1$ (overestimation of benefit).

2 *Type I failure: Investing effort in a disvaluable goal*

$$V(i) = 0, \ V(m) = \tfrac{1}{2}, \ V(f) = -1, \ \therefore \ V(i, m, f) = -\tfrac{3}{4}.$$

For utilitarianism, $V(i, m, f) = -1 - 0 = -1$ (overestimation of damage).

3 *Type II failure: Attaining valuable goals through disvaluable means*

$$V(i) = 0, \ V(m) = -1, \ V(f) = 1, \ \therefore \ V(i, m, f) = 0.$$

For utilitarianism $V(i, m, f) = 1 - 0 = 1$ (neglect of harmful effects of means).

Our postulate is of course normative not descriptive. According to it the badness of m may cancel out the goodness of f; hence an agent both rational and just will weigh his means as well as his goals. This point is of course of great importance in an ethical theory which, like ours, is based on an axiology. In our ethics, contrary to utilitiarianism and machiavellism, the end does not justify the means.

Our postulate is also relevant to action theory. Since there may be alternative means for attaining a given goal, we ought to choose the means optimizing the total value $V(i, m, f)$, rather than just the difference between the values of the initial and final states. Caution: *Optimization* is not the same as *maximization*. In many cases optima lie between minima and maxima. This is so because the maximization (or minimization) of any one variable is bound to alter at least one additional desirable feature in view of the fact that every variable (property) is related to at least one other variable. Examples: Maximizing individual liberty jeopardizes equality; maximizing equality imperils liberty; and mazimizing happiness puts both liberty and equality at risk.

Our assigning values to states and processes solves the puzzle that the nonexistence of something — e.g. the absence of a constraint — may have a (postive or negative) value. Thus, not being sick amounts to being in a state of good health, and not being at war is the same as being at peace. In general, the value of the absence of a state or process is the same as the value of the corresponding "positive" state of process.

Finally, we must answer an objection to the assigning of values to states rather than events or processes. Williams (1973 pp. 84 ff) has held that one often enjoys doing things, such as reading or rowing, for their own sake rather than for their consequences, e.g. learning or keeping healthy. This objection is invalid because doing something is a process, and every process is a sequence of states. (Recall Vol. 3, Ch. 5, Sect. 2.3.) So, if we value a process, we must also value the states that make it up. What is true is that utilitarians, the target of Williams's criticism, overlook the states that mediate between the initial and the final stages. This oversight, of which we have not been guilty, is mistaken because in some cases the intermediate states are the more rewarding, and in others we have the choice between different routes to a given goal state. This is why in real life one weighs the means as well as the goals. More on the means-end relation in Ch. 7, Sect. 1.5.

3.3 *Subjective Value (Utility)*

The word 'utility' is ambiguous: psychologists tend to interpret it as well-being, businessmen and economists as either value in use or as profit, and political scientists as social benefit. However, there is no doubt that in utility theory and its applications the term 'utility' is synonymous with subjective value or psychovalue as defined in Ch.1, Sect. 2.2. In fact, in that theory every person is attributed her own utility function(s), and the expected utility of the outcome of a course of action is calculated with the help of that function together with the subjective probability that the person assigns to the said outcome. In short, utility theory is subjectivistic, hence nonscientific. (Recall Ch. 2, Sect. 1.1.) This, together with its mathematical form and simplicity, may account for its popular appeal.

Since utilities or psychovalues are subjective, interpersonal utility comparisons are actually conducted in an intuitive rather than a scientific way. In this regard, utility is in the same boat with pleasure and pain, aesthetic quality and food taste. But, by the same token, it should be possible for physiological psychologists to come up with objective (physiological) measures of psychovalue. (Conjecture: The joint activity of the frontal lobes, which are the evaluation centers, and the limbic system, which is the "seat" of emotion, might yield such measure.) Should this breakthrough take place, utilities would be quantitated in a non-arbitrary way and they would become interpersonally comparable. But until such time we must make do with intuitive assessments. Hence the claims to the scientific status of all the disciplines that employ the concept of utility, in particular welfare economics and social choice theory, are unjustified. (See e.g. Hahn 1982, Eichner Ed. 1983.)

If utility (or psychovalue) were to be quantitated on a scientific basis, it should be possible to combine it with objective value. For example, it would be possible to set up a number of measures of the total (objective and subjective) value of an item, such as $W = \frac{1}{2}(U + V)$, $W = (U^2 + V^2)^{\frac{1}{2}}$ and $W = U + iV$, where $i = \sqrt{-1}$. It would also become possible to state in an exact manner the so-called law of diminishing marginal utility, or decelerated increment in subjective value — which must not be mistaken for the genuine law of diminishing returns, which occurs in economics. (A possible statement of the former law is Daniel Bernoulli's: The subjective value of an item is a logarithmic function of its objective value.) Consequently the alleged law could be put to rigorous empirical tests. And presumably it would utterly fail in such cases as the tycoon (who can never have enough of anything), investment in basic research and original design (the yields of which increase exponentially rather than logarithmically), and participation in collective endeavors.

But even if it became possible to quantitate utilities in a nonarbitrary manner it would make no sense to add up the utilities of a given item for a number of individuals, thus obtaining total (or social or inclusive) utilities. Indeed, utilities are *intensive* not extensive magnitudes: they are like densities and temperatures not like lengths and costs. The impossibility of adding the utilities of different individuals has been used by such sharp thinkers as Schlick (1930) and Sen (1985) to refute utilitarianism. However, Bentham's formula about the greatest happiness of the greatest number does not call for such summation: it says nothing about the total happiness of a social group. All it states is that we should procure that nearly every *individual* be given the chance of maximizing her well-being in her own way. The trouble with this principle in its normative version is that it is hard to maximize something that has not been quantitated in a satisfactory manner.

In short, current utility theory and its applications are nonscientific for failing to define the utility functions and to put their assumptions to empirical tests. Consequently all talk about utility, even when couched in mathematical terms, is so far largely preanalytical. Therefore we had better focus our attention on objective values, which have all the advantages of objectivity over subjectivity: they can sometimes be measured and compared in a scientific manner. However, one should not exclude the possibility that one day a rigorous concept of utility or subjective value will be proposed on the strength of psychological investigations. In the meantime all we have had from utility theory are entertaining paradoxes such as Prisoner's Dilemma and Newcomb's Problem.

3.4 *Value Calculi*

Many value calculi have been proposed. Some of them assume that all items can be evaluated in a quantitative (numerical) way, and that all value functions are additive. The first assumption was criticized in the preceding section; let us now examine the second. From the fact that the retail price (or exchange value) of two loaves of bread equals twice that of a single loaf, we tend to infer that all value functions are additive (or at least subadditive); i.e., that the value of a whole equals the sum of the values of its parts. (In obvious symbols, $V(x \dotplus y) = V(x) + V(y)$.)

Any number of counterexamples to the additivity assumption may be cited; let the following suffice. The biological value of water is not equal to the sum of the biological values of hydrogen and oxygen, if only because of the unique properties of water, such as its dissolving power. The social value of a city park does not equal the sum of the real estate value of all the square meters of which it is composed, if only because the arrangement of the trees and flower beds in it contribute to its total value. Again, the value of a well-assembled car is greater than the sum of the values of its parts. On the other hand the value of a criminal gang is even less than the sum of the (negative) values of its components. And occasionally meeting with two friends might be worse than meeting with only one, because the two may band together against oneself. In short, value functions are not generally additive.

Because value functions are not generally additive, no *general* calculus of value can be built. This impossibility is similar to that of building a general theory of analogy, or of measurement: in all such cases the nature of the case is of the essence. However, *special* value calculi, i.e. theories concerning particular species of items, such as commodities and actions, are in principle possible. For example, one may think of alternative (*vel*) and of joint (*et*) actions of a given kind, and may attempt to evaluate their outcomes in terms of the separate values of their constituents. A simple and cautious theory of this kind is the one compressed into the following three axioms.

Axiom 1 For all actions a and b, $V(a \text{ vel } b) \geqslant V(a), V(b)$.

Axiom 2 For any actions a and b,

$$V(a \, et \, b) \begin{cases} > V(a),\ V(b) \text{ iff } a \text{ and } b \text{ are mutually compatible} \\ < V(a),\ V(b) \text{ iff } a \text{ and } b \text{ are mutually incompatible.} \end{cases}$$

Axiom 3 For all actions a

(i) if $V(a) > 0$, then $V(\text{not-}a) \leqslant 0$;

(ii) if $V(a) < 0$, then $V(\text{not-}a) \geqslant 0$;

(iii) if $V(a) = 0$, then $V(\text{not-}a) = 0$,

where 'not-a' designates "not doing a".

These axioms are only semiquantitative. For instance, Axiom 1 does not tell us the worth of the alternative of two actions: it just sets a lower bound. This is as should be, for the value of any composite action depends on the nature of its constituents and on the specific mode of their combination. However, the point is that the above theory is not a *calculus* proper: it does not serve to calculate anything. Its only use is that of helping analyze the concept of an action.

3.5 *Summary*

We can grade items of a great many species but not all. For example, we do not rank remote stars or exotic plants, except perhaps aesthetically. And among the items that we do rank, some can be assigned quantitative values: this is the case with most goods and services.

In grading items of some sort we must distinguish between their objective and their subjective value, if only because an item may have the one but not the other. (For example, sunsets and junk food have only subjective values.) The objective value of an item may be equalled to the degree to which it meets or generates a basic need. In the former case the value will be positive, in the latter negative: Recall Figure 3.2.

When evaluating systems we must take their complexity and viability into account. However, this is not enough: the system must be good to begin with. Of two bad systems, the more complex and robust is the worse. When evaluating processes, in particular courses of action, we must take into account not only the total change but also the means by which it was brought about. A simple measure of the value of a process is the difference between the final and the initial values multiplied by a factor containing the value of the means or intermediate state(s).

The measurement of objective values poses methodological problems the solutions to which require substantive scientific or technological knowledge. Subjective values are usually deemed to pose no such problems: we assign them in an arbitrary fashion. The utility functions are not defined at all, and their values for real people are seldom checked. Hence utility theory and its applications are exercises in academic futility. They only produce an illusion of scientific respectability.

No general value calculus is possible because the value of a whole depends not only on the nature of its components but also on that of their links, hence on the value of the emergent properties of the whole. However, certain special value theories can be constructed, e.g. for actions of certain kinds. Still, these are not calculi proper because they do not allow one to compute the value of a compound item from the values of its constituents. This conclusion would have displeased Leibniz, who dreamt of replacing every disputation with a calculation. But it is a fact of life that, in matters of value, calculation may supplement argument, experience and sentiment, but not replace them.

PART II

MORALS

CHAPTER 4

ROOTS OF MORALS

When an animal values something it normally tries to get it, sometimes at the risk of its life. But other animals are likely to stand in its way, either because they occupy the same ecological niche — hence have the same needs — or because they are predators. Whenever this is the case there is either struggle (competition or predation) or cooperation. For example, male wolfs compete for females but cooperate in the hunt — and they do not kill conspecifics.

The higher gregarious vertebrates (mammals and birds) of the same species compete among themselves in some respects and cooperate in others. Usually they do so automatically, because they have inherited certain dispositions, or because they have learned such behavior patterns early in life. In any event the social behavior of such animals is constrained by internal inhibitions and external stimuli, e.g. threats. In the absence of such constraints the social group breaks up. For instance, male wolves fighting over females never fight to the death, but obey the tacit rule "Better defeated than dead"; and baboons and many other species cooperate in the defense of their groups against predators. Man is an exception, for in many societies he can occasionally engage in murderous competition and break some of the rules. But such antisocial behavior is not the norm and moreover it is learned, so it can be unlearned. Having the most plastic of all brains (hence behaviors as well), man is redeemable.

The constraints on human social behavior are made into more or less explicit rules or norms. These norms are learned, often the hard way, in the course of individual development. This is partly a process of enculturation, i.e. of learning (or internalizing) the socially acceptable behavior patterns. Being man-made, not natural laws, such rules or norms can be altered or even broken, as is the case with criminals and heroes. All the norms of human social behavior are kept, reformed, or repealed on the belief that they contribute to the well-being of some or all members of society: they are means to helping meet the needs and wants of members of society. Whether they do in fact help is a matter that can only be settled by an empirical study of the consequences of the actions governed by such rules. This statement reveals that our ethics is consequentialist.

Whereas some of the rules of human social behavior are moral, others are legal, and still others both moral and legal. For example, the condemnation of cruelty is moral, but there are some cruel laws; on the other hand the condemnation of rape is both moral and legal. In other words every culture, however primitive, favors two (or more) codes of social behavior: a moral code and a legal one; and the two, though distinct, partially overlap. Moreover, whereas moral codes emphasize duties, legal codes concern rights as well as duties. However, we shall argue that this asymmetry is irrational, and that in a good society rights and duties go hand in hand. In fact, the culmination of this volume will be a sort of Declaration of Universal Human Rights and Duties.

We shall be concerned mainly with moral rights and duties, not legal ones. The latter exist only within legal systems: There is no such thing as a legal right or duty outside a body of (positive) laws and jurisprudence in a society. Moral codes too are adopted or rejected by society, or at any rate by its dominant group. However, the rights and duties which are rooted to basic needs and legitimate wants — that is, to objective values — are far more fundamental and universal than any legal codes, bound as the latter are to political accidents and special interests. From a moral point of view (though not from a legal or historical one), moral norms take precedence over legal norms. After all, the whole point of any code of social behavior, as well as of any reform of it, is to ensure or block the exercise of moral rights and the fulfillment of moral duties.

In this chapter we shall explore the nature and sources of moral norms. We shall maintain that moral norms are ultimately rules of social behavior, for they come into play whenever self-interests (or rights) conflict with other-interests (or duties). Still, this does not entail that society is the only source of morality. Like values, morals have biological and psychological sources too, and this simply because the point of moral norms is to help realize (or inhibit) human values, which — as we saw in the preceding chapter — have all three sources. Thus we shall retain a grain of truth from each of the three non-theological schools concerning the origin of morals: biologism, psychologism, and sociologism. But we shall also add a dose of rationalism, since we can argue about moral problems and rules, and a dose of empiricism, since we care about the consequences of our actions (or inactions) and the performance of our moral principles.

1. RIGHTS AND DUTIES

1.1 *Right and Duty*

Human groups exist solely because every human being has needs and wants that can only be met with the help of others. If every individual were self-reliant from birth there would be no point in sociality, let alone in reciprocity or mutual help, and in the associated system of rights and duties made explicit in the moral and legal codes. But in fact no individual is self-reliant because nobody is complete, let alone perfect. We overcome our individual limitations only by joining others.

Sociality is a disposition and one that has been "encoded" in the genome of normal gregarious animals. The occasional social or antisocial individual does not last long and is unlikely to leave any offspring. Social selection puts out of action or even eliminates the asocial or antisocial persons, i.e. those who take without giving anything in return or who are unwilling to ask for help when in need.

However, like linguistic ability, sociality is only an inborn disposition. We must learn certain social skills in order to be accepted by other members of our social group(s), and even more so in order to induce others to alter traditional behavior patterns. Such social skills and graces are learned at home, in the neighborhood, at school, at work, and in informal social groups. Since the moral rules are among the norms of social coexistence, and since the latter emerge ultimately to secure the satisfaction of needs and wants, it follows that the ultimate source of morals are needs and wants, not only visceral but also psychological and social. Hence morality has the same sources as value. And, since every society evolves over time, it also follows that mores and morals coevolve with society — hence with needs and wants. Everything else, in particular rights and duties, is derivative.

Rights and duties may be partitioned into natural, moral, and legal. A *natural* right is one that an animal has conquered or been given. For example, a polar bear acquires by force the right to control a certain territory, and it gives her cubs the right to play only within her sight. Moreover the animal sees to it automatically, without any need for social control, that her offspring get the care they need: she thereby discharges her natural duty.

The concepts of natural right and duty become highly problematic in the case of humans because there are no human beings in the state of nature. Man is largely artifactual, i.e. a product of his upbringing and his activity in a given society. Change homes or societies, and different human beings will develop. This is why, unlike all the other animals, humans need a more or less explicit recognition of their moral and legal rights and duties. This is also why the concepts of right and of duty must be analyzed as high degree predicates: Animal v, when in state w, has the right (or duty) to perform (or refrain from performing) action x in society y when in state z. (For a different analysis see Kanger 1957.)

We start by elucidating the concepts of legal right and duty by stipulating

DEFINITION 4.1 If a is an action that a member b of a human group c is (physically) capable of performing, then

(i) b has the *legal right* to do a in $c =_{df} b$ is free to do or not to do a in c (i.e. b is neither compelled to do a nor prevented from doing a in c);

(ii) b has the *legal duty* to do a in $c =_{df} b$ has no right to refrain from doing a in c.

Note in the first place that a legal right is not the same as the effective exercise of it: ordinarily rights do not come with the means to exercise them. The same holds for duties: having the legal duty to do x does not guarantee that x will be done. Secondly, according to the above definitions rights and duties are vested in individuals because they concern actions performable by individuals. To put it negatively: There are no *collective* rights or duties, just as there is no social choice and no social action. The expression 'collective right (or duty)' must therefore be read as 'the right (or duty) of every member of a certain social group'. This is not to say that rights and duties are infrasocial (natural) or suprasocial (God-given). It just means that only individuals can be assigned rights and duties or deprived of them. The idea of collective rights and duties is a holistic myth that has been used to dilute personal responsibility and to discriminate in favor or against special groups.

The above definition tells us what rights and duties are but not which they are in a given society. The answer to the which-question is not philosophical but legal: it is found in the codes of law and the jurisprudence of a land.

In addition to legal rights and duties there are moral ones. The former are specified by the law, whereas the latter are accepted or self-imposed independently. There is no moral obligation to obey the law as such; moreover if a law is bad then it is our moral duty to fight it. Nor is every moral duty enshrined in a code of law. For example, there is no law ordering us to help one another. The most praiseworthy actions are those performed beyond the call of (legal) duty. However, in every society the legal and moral codes have a nonempty intersection. For instance, the moral rights to live and love, to work and learn, and many others, are enshrined in most codes of law.

In line with Chapter 1 we shall propose that a moral right is the ability to meet a basic need or a legitimate want, as characterized by Definitions 1.12 and 1.13. Likewise we shall propose that a requirable moral duty is a duty to help someone else exercise her legitimate moral rights. In value-theoretic terms: Moral rights are rights to realize primary, secondary or tertiary values (Definition 1.14). And requirable moral duties are duties to help others realize primary, secondary or tertiary values. The quaternary values, which are products of fancy or whim, remain beyond the ken of rights and duties, particularly in a poverty-stricken society.

What is the logical and methodological status of the above ideas on rights and duties? They are certainly not definitions, for these are supposed to hold come what may, and everyone knows that many contemporary societies do not recognize the right of everyone to well-being, let alone to reasonable happiness — except perhaps as an ideal far off in the future. Nor can our ideas on rights and duties be postulates, for they do not always match reality. There are two options left: to declare them natural rights and duties as stipulated in some theology, or to conceive of them as norms or rules.

The former move will not do in a science-oriented philosophical system like ours, and it will not help bring rights and duties down to earth. Besides, the thesis that rights and duties are natural is at variance with the finding of the social sciences, that morals are largely artificial and therefore subject to social change. The second alternative, namely to conceive of our ideas on rights and duties as norms or prescriptions for social behavior, has the advantage of helping us guide (in particular correct) social behavior, as well as redesign and rebuild society. Viewed in this way, i.e. as a system of norms, a morality turns out to be not only a sort of operating manual for handling the social machinery, but also a blueprint for building a better (or worse) society. (For a logic of normative systems see Alchourrón & Bulygin 1971.)

To put things more formally, we stipulate

NORM 4.1 If x is a human being in society y, and z is a thing or process in or out of x, then

(i) x has a *basic moral right* to z in y if and only if z contributes to the well-being of x without hindering anyone else in y from attaining or keeping items of the same kind as z;

(ii) x has a *secondary moral right* to z if and only if z contributes to the reasonable happiness of x without interfering with the exercise of the primary rights of anyone else in y.

According to this norm, everyone is entitled to enjoying well-being — even in miserable societies (Definition 1.18). Belief in such entitlement is instrumental in designing, proposing and implementing social reforms aiming at the effective exercise of basic human rights.

The above norm suggests that the pursuit of (reasonable) happiness is legitimate provided it does not hinder others from enjoying a state of well-being. But the norm does not state that everyone is entitled to reasonable happiness, because getting hold of the required means — which are often scarce — may involve entering into some competition, and in all competitions there are losers as well as winners. When the play is fair, those who win make use of special abilities which not everyone can acquire — as well as of a bit of luck, which does not strike everyone. (Luck = unexpected opportunity.) Of course, rational people will not persist along lines where they have no chance of winning honestly, but will try some alternative line. However, not everyone is rational, let alone lucky. Consequently even the best of societies are bound to contain some unhappy people: some because they have failed to realise all their ambitions, others because of neuroendocrine dysfunctions, still others because they have been victims of an education that stresses personal success above all.

Our next norm concerns duties:

NORM 4.2 If w and x are human beings in society y, and z is an action that w can perform (by herself or with help of others) without jeopardizing her own well-being, then:

(i) if x has a primary right in y to z or to an outcome of z, then w has the *primary moral duty* to do z for $x =_{df} w$ alone in y can help x exercise her primary moral right to z or to an outcome of z;

(ii) if x has a secondary right in y to z or to an outcome of z, then w has the *secondary moral duty* to do z for $x =_{df} w$ alone in y can help x exercise her secondary moral right to z or to an outcome of z.

Examples. Parents are normally the only people, or at least the best placed ones, to take good care of their children; hence they have the duty to do so. Likewise in most societies aged parents are dependent on their adult children for their well-being, whence the filial duty. However, many do not feel this duty to be an obligation.

Since scarce resources can be had, if at all, in limited quantities or in exceptional circumstances, some needs and wants can be met only partially. In other words, the effective exercise of rights and performance of duties is limited by the availability of resources and their distribution in each society. Note also that, whereas some duties can be delegated to others — such as nurses, servants or robots — most rights are not transferable except with adequate compensation. For example, a life prisoner may wish to be given the chance of freedom in exchange for his volunteering for a risky medical experiment. The key word here is 'volunteering'.

In manipulating rights and freedoms, particularly in trading rights for obligations, we should abide by

NORM 4.3 All of the basic moral rights and duties are inalienable except for trade-offs contracted between conscious and consenting adults under the supervision of a third party capable of having the contract observed.

Examples. The rights to live, work for compensation, choose one's friends, and have a say in matters of public interest, are morally inalienable. (Whence the immorality of military draft, unemployment, choosing other people's friends, and vote buying.) So are the duties to help our relatives, friends and colleagues, to respect other people's rights, and to preserve the environment. On the other hand property rights are typically alienable.

Caution: The alienable-inalienable distinction is not a dichotomy, for there is a grey zone composed of such rights as selling parts of one's own body, e.g. blood and a kidney. Another is the right to pollute, which is tacitly recognized by any laws or rules that impose fines on polluters without forcing them to stop dumping wastes. However, these particular cases, though legally grey, are morally black, for the sale of parts of one's own body puts life at risk and in the hands of the better off, and environmental pollution is bad for everyone.

Sometimes certain moral rights or duties are not enshrined in legal codes or, worse, they conflict with the law. For example, a political prisoner who has not committed any violent act has the moral right to try and escape from prison, and we all have the duty to denounce those who plan criminal actions, particularly aggressive wars. We generalize and adopt

NORM 4.4 Morality overrides the law of the land.

Right and duties can be grouped into five main categories: environmental, biopsychological, cultural, economic, and political: See Table 4.1. Although these are different, they are not mutually independent. For example, the free exercise of civic rights (e.g. voting) and duties (e.g. protecting public goods) is necessary to protect or facilitate the access to economic, cultural, biopsychological and environmental resources; and the enjoyment and protection of the environment are necessary for both sanitary and long-term economic reasons. Shorter: The moral rights and duties of all five kinds form a *system*.

The systemic character of moral rights and duties is not universally understood. On the one hand, liberals emphasize political and cultural rights, as well as certain economic rights, such as those of hiring and firing; on the other hand socialists emphasize the biopsychological rights and the rights to work, security and education. The former overlook the fact that it is hardly possible to play an active part in politics or to enjoy culture on an empty stomach; and the latter overlook the fact that only political liberty and participation can conquer or protect other rights. As for environmental rights and duties, liberals and socialists have traditionally ignored them — whence the rise of the "Green" party. To press for rights or duties of certain kinds while overlooking the rest is a mark of naïveté or worse. And to forget about duties altogether, or to put the defense of flag or party before interpersonal mutual help and international cooperation, is to invite jingoism and fanaticism, hence bloodshed.

It may be noted that the concept of a duty introduced by Definition 4.2 is much broader than that occurring in most contemporary ethical literature. Indeed, the latter focuses on comparatively trivial sacrifices,

TABLE 4.1. The five types of moral rights and duty and one example of each.

Type	Right	Duty
Environmental	Clean environment	Environmental protection
Biopsychological	Well-being	Help others enjoy well-being
Cultural	Learning	Teaching
Economic	Work	Workmanship
Political	Liberty	Popular participation

such as giving to charities, involving the anonymity of giver and taker. In our view supererogatory duties, particularly when having a narrow scope, are of peripheral importance. As well, the identity of giver and taker is often essential to moral action: I must take full responsibility for my duties as well as for the protection of your rights. And my children and friends ought to know whether I do my duty, so that they may take example or encourage me not to fail when the going is rough.

Finally, note that we began by attributing animals other than humans natural duties and even rights, but left open the question of animal morality. Moreover we restricted legal rights and duties to humans, because these seem to be the only animals capable of establishing, breaking, and altering conventions concerning social behavior. This entails that the animal rights activists, though well-intentioned, are mistaken in their belief that there is such thing as an animal legal right. However, there are *human duties* with regard to animals, e.g. to refrain from causing unnecessary pain. These duties are of a moral nature and some of them have been enshrined in regulations, particularly those regarding animal experimentation and the welfare of farm animals (protected by law in Sweden as of 1988).

1.2 *Rights Imply Duties*

Though neither of the norms proposed in the preceding section is deducible from one another, they are related in various ways. An obvious relation is this: A duty cannot be performed unless one has both the freedom and the means to exercise the right to do so. Other relations between rights and duties are far from obvious and they vary from one social group to the next. For example, the members of a pleasure-oriented group will tend to neglect their duties, whereas those of a stern religious or political group will tend to forego their secondary rights. We stipulate the following relations among the four items in question:

NORM 4.5 (i) Primary rights take precedence over secondary rights. (ii) Primary duties take precedence over secondary duties. (iii) Primary duties take precedence over secondary rights. (iv) An individual faced with a conflict between a right and a duty is morally free to choose either, subject only to condition (iii).

In other words, first things should come first, and survival of self and others first of all. In particular, the meeting of basic needs ought to dominate the satisfaction of wants; and among the latter the legitimate ones should take precedence over all others. Also, duty comes before pleasure. However (point (iv)), everybody has the right to sacrifice herself, even to the point of putting her well-being or even her life on the line. At the same time, nobody has the right to demand such sacrifice, or even to remonstrate anyone who fails to make it. In particular, nobody has the right to force or trick people into their death. More on this in Ch. 6, Sect. 1.2.

All three norms stated so far have been couched in the declarative, not in the imperative mode: they are propositions, not commandments. (However, they function as injunctions not as descriptions.) Being propositions, they can be conjoined together or with others to entail further propositions. A first consequence of Norms 4.1. and 4.2. is

THEOREM 4.1 Every right implies a duty.

Proof. Take a two person universe and call N_i a basic need, and R_i the corresponding right, of person i, with $i = 1, 2$. Likewise call D_{12} the duty of person 1 toward person 2 with regard to her need N_2, and D_{21} the right of person 2 toward individual 1 with regard to her need N_1. Our first two norms in the preceding Section can then be abbreviated to

Norm 4.1 $N_1 \Rightarrow R_1 \wedge N_2 \Rightarrow R_2$.

Norm 4.2 $[R_1 \Rightarrow (R_1 \Rightarrow D_{21})] \wedge [R_2 \Rightarrow (R_2 \Rightarrow D_{12})]$.

Using the principle of the hypothetical syllogism we obtain

$$[N_1 \Rightarrow (R_1 \Rightarrow D_{21})] \wedge [N_2 \Rightarrow (R_2 \Rightarrow D_{12})].$$

And, since everyone has basic needs, it finally follows, by modus ponens, that

$$R_1 \Rightarrow D_{21} \wedge R_2 \Rightarrow D_{12}. \qquad \text{Q.E.D.}$$

This theorem, and the postulate that all persons have certain moral rights (Norm 4.1), jointly entail

COROLLARY 4.1 Everyone has some duties.

We can extract further consequences if we adjoin further premises. We shall now add the anthropological, sociological and economic truism that no human being is fully self-reliant or self-sufficient:

LEMMA 4.1 Everybody needs the help of someone else to meet all of her basic needs and some of her wants (and thus realize all of her primary and secondary values as well as some of her ternary and quaternary ones).

This proposition entails the mutual help, reciprocity, or *quid pro quo* principle:

THEOREM 4.2 Helping implies being helped and conversely.

Proof. As in the proof of the previous theorem, imagine a two person universe. In this case Lemma 4.1 reads:

$$N_1 \Rightarrow H_{21} \wedge N_2 \Rightarrow H_{12},$$

where 'H_{21}' stands for "person 2 helps person 1 meet need N_1", and similarly for H_{12}. Suppose now that 1 helps 2, but 2 does not reciprocate, i.e. $H_{12} \wedge \neg H_{21}$. This is equivalent to the denial of $H_{12} \Rightarrow H_{21}$. Transposing Lemma 4.1 we get $\neg H_{21} \Rightarrow \neg N_1$. But N_1 is the case. Hence, by modus tollens, H_{21}, which contradicts our supposition. Therefore $H_{12} \Rightarrow H_{21}$. To prove the second conjunct we exchange the indices in every line of the preceding proof. Conjoining the two results we obtained the desired theorem, i.e.

$$H_{12} \Leftrightarrow H_{21}.$$

Life in a viable society requires a balanced mix of self-help with mutual help. This is why the most universal (least ethnocentric) moral maxims combine rights with duties. Witness the Golden Rule. Any moral doctrine that, like hedonism, admits only rights, or — like Kantianism — admits only duties, may be said to be *unbalanced*, hence unviable. Only psychopaths, professional criminals, warmongers, robber barons, tyrants, and members of societies in the process of disintegration, refuse to engage in mutual help — to their own long-term detriment and that of their society. (For the centrality of mutual help in animal societies see Kropotkin 1902.)

Normal social behavior is prosocial (Definition 1.10) — i.e. a matter of give and take, of balancing rights with duties and vice versa. For example, every able-bodied person has the right to work and the duty to work well; and every citizen has the right to have a say in the manner he is being governed, and the duty to participate in the political process — if only to see to it that his rights will be respected.

Although basic needs and legitimate wants generate rights, and these duties, the latter in turn restrict rights. In particular, every one of my liberties ends where someone else's begins: Your right is my duty, and your duty is my right. (See Hobbes 1651 p. 67.) The law of reciprocity, or give and take, is nothing more and nothing less than a tradeoff between rights and duties. More precisely, there is a negative feedback between these: See Figure 4.1.

The exclusive pursuit of one's own interests with disregard for the well-being of others is not only socially destructive. It is also chimerical, for nobody is fully self-reliant, and a person does not get help from others on a regular basis unless she is willing to do something for them. (Even slavery and serfdom involved trade-offs, in particular the exchange of labor for security.) As for the exclusive dedication to the welfare of others, without regard for one's own well-being, it is self-destructive and therefore inefficient. A sick saint is as useless to society as a sated pig.

All of the preceding can be summarized into a single moral norm that we shall make into the supreme principle of our morality:

NORM 4.6 *Enjoy life and help live.*

The first conjunct is of course the devise of hedonism or individualistic utilitarianism. It is tempered by the second conjunct, the lemma of philanthropism. (The latter must not be mistaken for the negative utilitarian maxim "Let live" or "Do no harm", which goes well with the egoistic maxim "Do no become concerned, do not get involved".) Taken together the two parts of Norm 4.6 synthesize egoism and altruism, self-interest and other-interest, egocentrism and sociocentrism, autonomy and heteronomy. It may therefore be called the *selftuist* principle. We submit that this principle, far from being just a pious

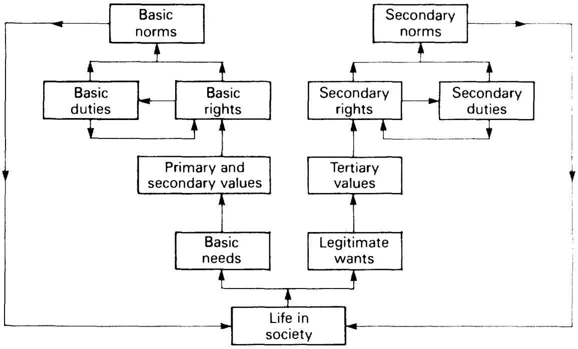


Fig. 4.1. Needs and wants of people living in society as the ultimate sources of rights and duties. Every right generates at least one duty and conversely. In turn, rights and duties are enshrined in moral norms, which react on life in society.

thought, is the most effective guide for the design or redesign, building or maintenance of a good society, namely one in which everyone has the chance of becoming reasonably happy (according to Definition 2.3).

Lest Norm 4.6 be misunderstood as Voltaire's Candide's invitation to cultivate one's own garden, we shall explicitly add its correlate in public morality, namely

NORM 4.7 *Seek the survival of humankind.*

Actually the two norms are equivalent, because we will not help the survival of humankind unless we look after ourselves and others, and if we do this we shall make a contribution to the survival of the species.

The preceding norm is the moral counterpart of Norm 2.2, which asserted that the survival of the human species is the supreme good for any rational person. Consequently we may adopt Norm 4.7 as the highest principle of our moral code.

Finally, note that rights and duties can be *positive* or *negative.* Positive rights (mays) are permissions to do, and negative rights (need nots) permissions not to do. Similarly positive duties are obligations to do (do's), and negative duties obligations not to do (don'ts). Table 4.2 exhibits a few examples of moral maxims in every one of these categories, all of them compatible with Norms 4.6 and 4.7.

1.3 *Summary*

The block diagram in Figure 4.1 summarizes our view on the sources of morals. The ultimate wellspring of morality is life in society, which has

TABLE 4.2. Instances of positive (+) and negative (−) rights and duties.

Sign	Rights	Duties
+	Seek self-realization. Participate in social activities Take whatever pleasures do not harm self or others.	Support yourself and dependents. Participate in social activities. Care (fraternity) and share (equality).
−	Don't miss opportunities Resist exploitation and oppression. Do not feel bound to do anything beyond the call of duty.	Do not use people. Avoid violence. Do not do to others what you would not like done to you.

basic requirements and generates wants. Needs and desires generate values, and these are the foundation of rights. But in any viable society the exercise of a right imposes some duties, which in turn constrain the former. Finally, rights and duties are the subject of rules of various kinds, in particular moral norms. In turn, these regulate social behavior and thus react back on life in society. So much for structure.

The content of each block in the diagram is determined by the kind of life people live in their society. However, the basic human needs are, by definition, common to all human beings. Hence the corresponding rights and duties are universal or cross-cultural. These are the ones enshrined in the basic moral norms. One of them is the principle of reciprocity or mutual help. However, this norm is insufficient, for two persons may help one another in committing a crime or in committing suicide. Hence the need for a higher principle: *Enjoy life and help live.* This maxim combines the need for self-preservation and the desire of self-realization with the duty (and convenience) of assisting fellow humans in achieving similar goals.

2. MORALS

2.1 *Right and Wrong Actions*

The notions of moral right and duty elucidated in the preceding sections allow us to introduce those of morally right or wrong actions, as well as that of a moral agent. Here is the first convention:

DEFINITION 4.2 An intentional human action a is

(i) *morally right* $=_{df}$ a facilitates the exercise of a moral right or the fulfilling of a moral duty of either the agent or someone else;

(ii) *morally wrong* $=_{df}$ a hinders the exercise of a moral right or the fulfilling of a moral duty of either the agent or someone else;

(iii) *morally neutral* $=_{df}$ a is neither morally right nor morally wrong.

Examples. It is morally right to enjoy well-being and to help others improve their well-being. It is morally wrong to exploit or oppress people, because exploitation and oppression restrict the ability to enjoy well-being. It is morally neutral to contemplate a landscape or to scratch one's head. And unintentional actions, regardless of their outcomes, are neither morally right nor morally wrong.

Rightfulness and wrongfulness come in degrees. Thus whereas some actions only facilitate the exercise of a right or the fulfilling of a duty, others do much more than this: they are supererogatory actions that may involve sacrifice on the part of the agent. Likewise wrongfulness ranges between venial and cardinal sin.

We have not defined "right" in terms of "good" because one can think of any number of well-intentioned but foolhardy or incompetent, hence inefficient or even counterproductive deeds. Thus the concepts of good and right are logically independent from one another. However, it might be possible to defend the following alternative

DEFINITION An action a is *morally right* $=_{df}$ a is the best among all the possible good actions to the same end, i.e. that employing the most suitable and least harmful means, and having the least noxious side effects.

The problem with this definition is that in real life it is hard if not impossible to pick up the best of anything, particularly when in a hurry or when the various options, far from being at hand at the same time, appear sequentially and unpredictably. For the latter reason an important school of thought in management science recommends looking for satisficing rather than maximizing solutions (March & Simon 1958).

We judge the moral worth of an action by the intention of its doer rather than by its outcome. Occasionally a good action has an unintended bad consequence, whereas an evil one may have unintended good results. The cause of such unintended consequences is that the outcome of an action depends on three factors: the agent's skill, the state of the target, and the state of its environment (in particular the means available to the agent). We often lack the necessary skill or knowledge to bring about the desired change in the state of the object of our action. And even when in possession of both skill and knowledge, the occurrence of unpredictable (in particular random) events may spoil the best designed course of action.

Note that our analysis is consequentialist, not deontological: it enjoins us to study the situtation, as well as our abilities, and assess the possible consequences of our actions before taking action, instead of applying mechanically some received norms. However, ours is not a utilitarian analysis for, unlike Bentham (1789) and his followers, we do not define a right action as one causing pleasure or having a positive utility. In fact, the fulfilling of a moral duty may have disagreeable consequences. (Kant went to the extreme of claiming that doing one's duty is always painful, which is neither true nor conducive to moralization.)

Note also that our analysis or, for that matter, any analysis of the notions of right and wrong, is unacceptable to the moral intuitionists, such as Sidgwick (1907), Moore (1903), Scheler (1913), and Ross (1930), for whom those are simple ideas, i.e. unanalyzable ones. We submit that intuitionism, whether in moral philosophy or in epistemology, is a variety of irrationalism and as such incapable of helping us face reality with the help of reason and experience (Bunge 1962b). More in Ch. 7, Sect. 1.5.

The previous definition allows us to introduce

DEFINITION 4.3 An animal *b* is a *moral agent* $=_{df}$

(i) *b* knows right from wrong;

(b) *b* approves or disapproves of right and wrong actions, i.e. takes sides; and

(c) *b* acts on the strength of his own approval or disapproval of right and wrong actions.

Examples. Any normal child above two years knows the reciprocity rule and has a strong sense of fairness. Therefore he is a moral agent, albeit in limited domains (Piaget 1932). Mentally retarded and severely sick or senile people may not be able to distinguish right from wrong, whence they are not moral agents. It is an open question whether animals can tell right from wrong or whether, when they do their duty, they behave, because they have been conditioned, without ever having moral doubts or qualms. As for computers and robots, they are at most moral agents by proxy, for whatever decisions they make derive from instructions received from their programmers. (Recall Vol. 7, Ch. 5, Sect. 3.1.)

Note the following points about the preceding definition. Firstly, being a moral agent involves having some moral knowledge, however crude. Secondly, the moral agent cannot be impartial. But she must make an effort to be (*a*) *objective*, i.e. to examine both sides of a moral conflict and weigh the merits and demerits of every side; (*b*) *fair*, i.e. respect the rights of all the parties in conflict, and (*c*) *unselfish*, i.e. not seek personal advantage at the price of someone's well-being. In short, the essence of the moral viewpoint is not impartiality but objectivity, fairness and unselfishness. Thirdly, the moral agent cannot be passive out of moral blindness, cowardice, or weakness of the will. We tend to fight the wicked but to forgive the morally weak although the latter may cause more harm than the former. This attitude is mistaken, because the devils would be unable to do their job without the passive complicity of morally spineless people.

Moral and legal codes concern exclusively moral agents, and they all include some notion or other of moral responsibility. We characterize the latter by

DEFINITION 4.4 A person *a* is *morally responsible* for action *b*, or for action *b* not having been committed $=_{df}$

(i) *a* is a moral agent;

(ii) *a* is fully conscious of the intentions that triggered *b*, or of those that blocked the performance of *b*;

(iii) *a* a does (or orders somebody else, or instructs a nonhuman animal, or a machine) to do *b*, or to refrain from doing *b*, of his own free will (i.e. not under compulsion).

By virtue of the first clause all and only moral agents are morally responsible for their own actions or those that they instigate. Accordingly people acting under the influence of drugs forced on them cannot be held morally responsible for what they do or refrain from doing. This does not exempt those who act or fail to act under the influence of drugs taken willingly; in particular it does not exempt drunken drivers. Nor does it excuse people who torture, maim or kill non-combatants, or political prisoners, or prisoners of war, under orders from their superiors. Indeed, they could attempt to desert or, even better, avoid becoming involved in atrocities by declaring their conscientious objection to bearing arms or to serving in a repressive police force. After all, facing moral problems and taking the responsibility for our actions and non-actions is part of being human.

According to our definition we are morally responsible not only for our intentional or deliberate actions but also for faults of omission, such as negligence or failure to act at the right moment. Whoever is in charge or control of an event the outcome of which is beneficial or harmful to others, is responsible for that event or for the failure to trigger it.

According to Definition 4.4 only individuals can be morally responsbile, for the simple reason that only individuals can have a conscience. To put it negatively: there is no such thing as collective moral responsibility. All there can be, is the sharing of all the members of a group in a given responsibility. Therefore collective reprisals are moral monstrosities.

As well, it is mistaken to shift responsibility from the individual to society ("the system", "the establishment"), even though it is true that a bad social organization breeds some ill-behaved individuals. We may *explain* X's wrongdoing in terms of X's living in a defective society, but we cannot morally *justify* X's behavior, for X is responsible for whatever X does or fails to do. Unless we admit this, social and moral reform become pointless.

As with responsibility, so with credit and blame: these too can only be vested in individuals, as stipulated by

DEFINITION 4.5 A person a is to be *credited* with (or *blamed* for) action $b =_{df}$

(i) a is morally responsible for b, and

(ii) b is morally right (or wrong).

A merit of this definition is that it is far shorter and clearer than any of the abstruse theological disquisitions on blame; it is also more useful than the ordinary language analyses of the various significations of the word 'blame'.

Responsibility and blameworthiness are necessary but not sufficient to secure a moral order or to prompt a progressive moral reform. Many a tyrant has proudly taken the responsibility for his crimes, yet remained unpunished for them. We also need accountability to those likely to be affected by one's actions, or those elected or appointed to see that antisocial behavior be prevented or corrected. Therefore we add

NORM 4.8 All moral agents are accountable for those of their actions and failures to act that affect others.

This norm tacitly denies that anyone has the duty to account for his strictly private affairs: As long as his behavior does not harm anyone else, nobody has the right to pry into his affairs, let alone to reprehend or punish him. The same norm stresses that all responsibility is personal, never collective and therefore diluted. Where no individual takes credit or blame for what he does or fails to do, disaster is bound to occur. For example, it appears that the Chernobyl nuclear accident (1986) was due to the system of "collective responsibility", whereby nobody in particular was able to make the right decisions in time.

It will be noted that we have been dealing with moral not legal responsibility. Legal responsibility is defined by law and contract: it does not occur outside a legal context. Consequently legal positivists and contractarians are not concerned with moral responsibility — but everyone else is. For example, although the parents, friends and teachers of a professional adult criminal are not legally responsible for the latter's misdeeds, they may be morally responsible, for having failed to give him the proper moral education or for not having helped him acquire the skills necessary to engage in an honorable gainful occupation.

2.2 *Moral Problems*

We know what problems in general are (Vol. 5, Ch. 7, Sect. 4). Moral problems are a very special kind of practical problem: they can only be recognized by animals endowed with a moral conscience, and solved by those having special skills and knowledge. We characterize a moral problem through

DEFINITION 4.5 A moral problem is a conflict between either a person's rights and her duties, or between her duties, or between the rights of two different persons.

Example 1. A person has promised to marry another but eventually finds out that he is no longer in love with his betrothed. Should he keep his promise even at the risk of making at least two people unhappy?

Example 2. A doctor treats a terminally ill patient. Should he tell him the truth even at the risk of increasing his patient's misery and hence abbreviating his life?

Because rights and duties are ruled by norms, the preceding definition entails

COROLLARY 4.2 Every moral problem is a norms conflict.

This suggests that a way of going about solving moral problems is to rank moral norms, much as we ranked moral rights and duties in Norm 4.2, opting always for the dominant norm. More on this in the next Section.

Because the exercise of rights and the fulfillment of duties amount to the realization of values via the meeting of needs and wants, we infer

COROLLARY 4.3 All moral problems are value conflicts.

The converse of this proposition is false. For example, someone may be torn between wishing to spend her last dollars on a meal, or on materials to finish a painting that may or may not fetch her the price of the meal. For a value conflict to be a moral problem at least two people, or a human being and an animal, must be involved; and at least one of them must need the help of the other or must compete with her for something or someone.

Corollary 4.3 suggests a second way of solving moral problems, namely ranking values belonging to the same category (Ch. 1, Sect. 3.1) and opting for the higher value. Thus in the case of the physician faced with the terminally ill patient, his task is comparatively easy, for to him truth has mainly an instrumental value, the well-being of his patients being his top value. Other problems are of course harder. One of them is the conflict parents face between parental love and the parents's duty to educate their children. They must avoid the extremes of alienating severity and corrupting permissiveness. This case calls for a compromise involving concessions on both parts, constrained only by certain basic moral rules.

Since all moral problems are norm or value conflicts, and since conflict resolution requires some effort in addition to good will and knowledge, any solution to a moral problem, however beneficial it may be to the parties involved, carries some cost or loss to at least one of them. In this regard moral agents should behave much like managers though with different intentions: they should perform some cost-benefit analysis, however crude, before plunging into action regardless of its consequences. Only fools and prigs can afford to ignore costs to selves and others. In this regard utilitarianism is right: If we wish the good and the right to prevail on balance in the end, we must be somewhat calculating (though not selfish), and we must be prepared to pay the price, which includes the stress caused by the action and by the uncertainty of its outcome.

All moral problems boil down to identifying and resolving conflicts of interests, in particular self-interest vs. other-interest. Before meeting Friday, Robinson Crusoe faced no moral problems on "his" island. (However, a modern day Robinson would feel obliged not to pollute his environment unnecessarily in the expectation that others might arrive in future.) Therefore we cannot accept Dewey's dictum about "the fundamental moral nature of the social problems". Rather, we face moral problems because (*a*) we live in society, (*b*) we need others for our own survival and others need us for their own, (*c*) we can help or harm others, (*d*) we have options or can produce them, (*e*) we can get to know right from wrong, and (*f*) we are free, within bounds and at least sometimes, to choose what we judge to be morally right. (For the last three points see Simpson 1964.)

For these reasons we make the following

POSTULATE 4.1 All moral problems are social problems, i.e. either problems concerning personal behavior to others, or problems concerning the preservation or renovation of society.

But, although moral problems are engendered by social life, not conversely, their solution involves moral principles, because such solutions are bound to benefit some people while harming, or at least not benefiting, others. Hence we also need

POSTULATE 4.2 All solutions to social (economic, political or cultural) problems presuppose moral principles.

Postulate 4.1 involves a distinction between personal and social moral problems, hence moralities. And Postulate 4.2 entails that sociotechnology, unlike basic social science, cannot be morally neutral. In particular, whereas the description of an economy is morally neutral, a prescription for an economy involves a moral *parti pris*.

On the one hand personal moral problems are the exclusive responsibility of certain individuals, e.g. as parents or teachers. On the other hand social moral problems are the responsibility of all the adult members of a social group. There are then two kinds of morality, personal and social, and one of the tasks of the moral philosopher is to try and render them mutually compatible. Still, since only individuals can be moral agents, it behooves them, not collectivities, to abide by a social as well as personal moral code. (More on private and social morals in Ch. 6.)

Now, moral rights and duties, hence moral problems, may concern items of five types: environmental, biological, cultural, economic, and political (Section 1.2). Consequently we may distinguish the following types of moral problem, whether personal or social:

(1) *environmental* (e.g. pollution and the depletion of non-renewable natural resources);

(2) *biological* (e.g. health care and birth control);

(3) *cultural* (e.g. the control of public information channels and of junk culture);

(4) *economic* (e.g. economic inequality and consumer protection);

(5) *political* (e.g. the arms race and power abuse).

This is a typology not a classification, for every one of the above sets overlaps partially with some other set or sets. For example, the arms race deteriorates the environment and degrades the economy, which in turn causes cultural decline as well as a decline in the quality of health care services.

The point of the above typology is to remind us that moral problems can be very complex, far more so than that of choosing to behave like a good Samaritan or like a passive onlooker of human misery. The solution to some of the above mentioned problems requires more than moral sensibility, social skills, common sense, and ordinary knowledge. It may also call for a good deal of scientific and technological (in particular sociotechnological) knowledge and political action. Just think of designing any efficient social program.

The traditional societies, which evolved slowly, did not face most of the above-mentioned problems, particularly those requiring modern technology. Industrial civilizations change so quickly that unprecedented moral problems keep cropping up. A fanatic who sticks to some fixed set of moral norms designed for a traditional society is incapable of detecting, let alone solving, any of the moral problems raised by the nuclear arms race, overproduction, or overpopulation. He will tend to ignore them, focusing instead on side-issues, or he may wish to tackle them as if they were old problems, only somewhat more complex. For example, he will think of nuclear weapons as ordinary weapons, only more powerful, rather than as the first arms capable of wiping out the entire human species.

In addition to facing or skirting moral problems of different types we may tackle or elude moral problems of different orders of magnitude. We distinguish the following:

(i) *first order of magnitude*: global problems, i.e. problems affecting the whole of mankind, such as what to do about the arms race, overpopulation, the depletion of non-renewable resources, and environmental degradation. These are moral problems because they concern the survival of humankind. But they are also political and technological problems, for they can only be solved through political measures taken in the light of the best available technology;

(ii) *second order of magnitude*: social problems affecting large social groups, often as many as half of the world population, such as poverty, bad health, unemployment, state violence, lack of political rights, illiteracy, and the neglect of basic duties out of inability to fulfill them;

(iii) *third order of magnitude*: individual problems such as whether to steal in order to feed one's family, to obey an army draft order, to perform an abortion, or to help a neighbor in distress at one's own risk.

Regardless of type and order of magnitude, every moral problem is a practical social problem — but the converse is false. Moreover all moral problems are of the disjunctive form "*A vel non-A*", where *A* is a right action. Being of a practical nature, moral problems are not solved by consulting moral codes (particularly if obsolete) or works in moral philosophy. Codes serve to identify problems and books to analyze them; at most, codes and books may offer broad guidelines. Moral problems can only be solved correctly by adding some practical knowledge to our moral sensibility and reasoning. For example, once we have decided to help the victim of an accident, our problem is how to do it in the most efficient way. When the time for action comes, the moral problem has turned into a technical one.

A good piece of advice concerning moral and, in general, practical problems, is never to aim for perfect solutions. Perfectionism is to be shunned in the practical sphere because practical problems are usually pressing and often very complex. A useful first approximation attainable now is usually preferable to a better solution attainable with the help of more powerful means likely to be obtained later, for in the meantime the people affected by the problem may be deprived of their means of livelihood or even of their lives. In practice perfectionism is paralyzing. Therefore it is sometimes used as an excuse for inaction.

Are all moral problems soluble or is it the case that, like in mathematics, some of them are inherently unsolvable? Most moral philosophers are optimistic. In particular, act utilitarians believe that the best solution is easily found by estimating (subjective) probabilities and (subjective) values, whereas game theorists believe that it is all a matter of writing down payoff matrices with invented entries. The truth is that, just as there is no single and fool-proof method for solving all engineering problems, so there is no single and fool-proof procedure for solving all moral problems. The reason is simply that the method of effective solution depends on the specific nature of the problem, which may involve tough conceptual (in particular moral) and practical (in particular political) items. (See also Hampshire 1983 p. 152.)

Some moral problems are insoluble because they call for resources that cannot be made available in time. Others are soluble but do not get tackled because of lack of good will (or even due to moral blindness). The latter is the case with many of the first and second order magnitude problems mentioned above. On the whole, then, given good will, most moral problems are soluble, at least in principle and given time. But any efficient and lasting solutions to them require specialized, in particular, technological knowledge. We shall return to this point in Part IV on action.

2.3 *Morals*

A moral (or morality, or moral code) is a set of norms stipulating what is right and what is wrong, together with an order of precedence or dominance on that set. More precisely, we propose

DEFINITION 4.6

(i) A *moral norm* is a rule stipulating that actions of a certain type are morally right or wrong (according to Definition 4.2);

(ii) A moral norm is *right* if and only if it recommends performing a right action. Otherwise the norm is *wrong*;

(iii) a *personal* (or *private*) *morality* (or *moral code*) $\mathcal{M}_p = \langle M_p, \gtrsim \rangle$ is a set M_p of moral norms governing private life and inter-personal relations, together with a priority or dominance relation $\gtrsim$;

(iv) *a social* (or *public*) *morality* (or *moral code*) $\mathcal{M}_s = \langle M_s, \gtrsim \rangle$ is a set M_s of moral norms governing the relations between individuals and social systems (in particular governments) or their representatives, together with a priority or dominance relation $\gtrsim$;

(v) a *moral* (or *morality*, or *moral code*) $\mathcal{M} = \langle M_p \cup M_s, \gtrsim \rangle$ is the union of a set of personal moral norms with another of social moral norms, together with a relation $\gtrsim$ of priority or dominance;

(vi) a *metanorm* is a statement of the form "$R_i > R_j$", where R_i and R_j are rules in the union (or logical sum) of the sets M_p and M_s;

(vii) a *metamoral* (or *ethical*) statement is a statement of the form "Moral code A is superior (or preferable) to moral code B for social group C if goal D is to be achieved for the benefit of the members of C".

The third clause of the preceding definition deals with personal morality whereas the fourth concerns social morality. How are these two related? According to individualism, public morality ought to be subordinated to private morals: the former would be just a means to protect individual rights — particularly of those who have the "initial endowments" required to exercise them. (Duties would not matter unless explicitly stipulated in contracts.) On the other hand holists claim that private morality ought to be subordinated to public morality: the former would be just a means to have public duties observed — particularly when the public interest is defined as that of the ruling class. (Rights do not matter unless graciously granted by society or its rulers.) A systemic philosophy like ours (Vol. 4), containing a moral philosophy that holds rights and duties to come in pairs, countenances neither individualism nor holism. Nor can it subscribe to the subordination of duties to rights (individualism) or conversely (holism).

In our view the morally right relation between private and public morality is one of *mutual adjustment*, not one-sided subordination. For example, if we want to protect the right to life we have the duty to participate in public affairs so that our governments prepare for peace and protect the environment — without which humankind will not survive. In turn, governments that do protect effectively our individual rights and do something for the survival of humankind are entitled to our support, e.g. in observing the (just) laws of the land and paying (justified) taxes. More precisely, we propose the following metamoral (or ethical) rule:

NORM 4.9 (i) The ultimate goal of both private and public morals is the survival of humankind. (ii) The only morally right means that may be employed to secure the survival of humankind are (*a*) the protection of individual rights and (*b*) the observance of the individual duties paired off to those rights.

Some moral philosophers claim that what is typical of moral discourse is the occurrence of the word 'ought' (or 'obligation') in it. This view, perhaps unavoidable in the perspective of linguistic philosophy and against the foil of traditional authoritarian morals, is mistaken on several counts. Firstly, the norms concerning moral rights constitute a clear counterexample: indeed, they involve permissions not obligations. Another is the collection of nonmoral rules of everyday life, such as "Thou shalt daily milk thy cow". A third is the collection of technological rules, such as "You ought to aim for optimal output/input ratios". Secondly, in principle it is always possible, and clearly advantageous from a logical viewpoint, to state moral norms is a declarative mode, as we have done in the preceding with a single exception.

Every moral norm and, a fortiori, every moral code, refers to (*a*) the moral agent supposed to abide by the norm, and (*b*) whatever is likely to be affected by the action(s) ruled by the norm. These, the targets of the agent's action, can be other animals (human or not), social groups (e.g. governments), artifacts, ecosystems, etc. The inclusion of artifacts among the possible referents of moral norms should raise no eyebrows, for we are supposed to keep in good repair the useful ones (so that others may use them, or so that no resources be wasted), and destroy or recycle the noxious ones. Likewise, the inclusion of ecosystems should not be problematic, for their deterioration is bound to cause pain or even loss of life. (Recall that in Sect. 1.1 we mentioned duties concerning the environment.)

The members of M_p and M_s are rules about actions, but any statement of the form "$R_i > R_j$", where R_i and R_j are contained in M_p or in M_s, is about norms, hence it is a metanorm. And the formulation of a moral problem is a disjunction of the form "$R_i > R_j$ or $R_j > R_i$", so that it contains metanorms although it is not itself a metanorm.

Note that a (total) morality is not merely the sum of a personal and a social morality because, in addition to the rankings within M_p and M_s, it contains preference statements involving norms of the two kinds. Thus, the morality shared by most of the members of any given social group contains rules to the effect that certain norms of social behavior take precedence over certain rules of private behavior.

Finally, note that not every norm in a socially accepted moral code may be right: recall clause (ii) of Definition 4.6. For example, if we admit Norm 4.6 we shall conclude that it is morally wrong to kill, except perhaps for mercy or in self-defense. Hence any moral code that condones killing for any other reasons, even the killing of murderers, will be regarded as morally wrong. This example suggests that, once we have adopted a set of basic moral norms, such as our Norm 4.6, we may pass moral judgment over any moral code. However the examination of actual moral codes in the light of moral (and rational and empirical) principles is the business of ethics as distinct from morals. We shall come back to ethics in Chs. 7 to 9.

2.4 *Summary*

Morality is concerned with intentional or deliberate human actions. Actions can be morally right, wrong, or neutral, depending on whether they facilitate, hinder, or do not affect the exercise of moral rights or the fulfilling of moral duties.

An animal is said to be a moral agent just in case it is capable of telling right from wrong. All and only moral agents are responsible and accountable for their actions. There are no such things as collective moral agents and collective responsibility: agency and responsibility are individual.

All moral agents face or elude moral problems. A moral problem is a conflict between values or norms. Being responsible for his actions, the moral agent is expected to seek adequate solutions to his moral problems. In so doing he cannot proceed on his own because every moral problem is ultimately a social problem. Nor can he solve any social problems in a morally neutral fashion, because every solution to a social problem is bound to favor or harm some people more than others.

Moral problems are of several kinds — personal and social, environmental and cultural, big and small, and so on. Hence there can be no ready-made and universal moral problem solver. Moral codes help spot moral problems and guide rational thinking about them, but such problems, if important, must be investigated responsibly the way engineering or medical problems are. That is, their solution requires substantive knowledge relevant to the nature of the particular case, and some such knowledge may not be available but may have to be produced. I.e., moral problem solving may require research.

A moral code is an ordered system of norms specifying what is right and what is wrong. While some such norms regulate interpersonal activities, others guide the behavior of individuals *vis-à-vis* social groups or their representatives. Every moral code is supplemented with metamoral (or ethical) norms stating that such and such norms are superior to such and such other norms. The supreme metamoral norm in our system is that our main concern should be the survival of humankind, and that the only legitimate means to pursue this desideratum involve the protection of individual rights and the observance of the corresponding duties.

3. SOURCES AND FUNCTIONS OF MORALS

3.1 *Biological*

A moral philosophy which, like ours, is based ultimately on the recognition of needs and wants, must admit the biological roots of some moral values, hence of some rights and duties, whence of some moral norms as well: See Figure 4.1.

It would seem that there are no *natural* rights: that all rights are social, both because they concern social behavior and they are conquered, bestowed, or withdrawn by people acting together. On the other hand it may be argued that some *duties* "are in the genes". Thus the bird that feeds its chicks is genetically "programmed" to do so: It performs its parental duties without any need for a moral conscience or for social vigilance. It is possible that some human duties have a similar genetic root; in particular this may hold for child care and the impulse to help others. But it is obvious that such natural proclivities can be strengthened or weakened by social circumstances. It is also evident that other duties, e.g. that of respecting public property, are just as learned, hence social, as the rights that generate them. To state that "every moral, pressure or aspiration is essentially biological" (Bergson 1932 p. 103) is a gross exaggeration. (For further examples of ethical biologism see e.g. Ebling Ed. 1969, and Stent Ed. 1980.) In general, biological determinism is inadequate, for exhibiting only one side of human nature.

If some of the roots of morality are biological, then (*a*) other animals, particularly those closely akin to us and hence with similar basic visceral needs, must show signs of moral behavior, and (*b*) the moralities prevailing in different human societies must have a nonempty intersection: i.e. there must be some universal moral principles concerning the preservation of the individual, his kin and his neighbor.

There is some empirical evidence for these two theses. But regrettably ethologists, anthropologists, social psychologists and moral psychologists have been somewhat remiss in their duty to amass the evidence. The thesis of animal morality, emphatically denied by theologians and idealist philosophers, has been confirmed by uncounted cases of respect for the life of conspecifics, empathy, and mutual help. Even if all such behavior patterns turned out to be genetically determined, the moral character of the corresponding norms would not be disproved for, after all, morality is about how to live in society.

Holding that morality has biological roots is not the same as stating that these are the only roots, let alone that we are born good or evil, altruistic or selfish, cooperative or competitive. The innatist thesis is false, for the newly born human is incapable of deliberate social behavior, if only for lack of motor coordination. The immature state of his brain, as well as the way he gradually learns (internalizes) the prevailing moral rules, and moves from egocentrism to cooperation, suggest that at birth humans are neither good nor evil. We have only an inborn potential to become either through education and life experience — or rather to become a mixture of goodness and wickedness, as R. L. Stevenson would have said. The same holds for knowledge and, in particular, for linguistic knowledge. (Recall Vol. 5, Ch. 1, Sect. 3.1.)

As for the thesis of the existence of a universal morality, i.e. a core shared by all moral codes and rooted to human nature, it is suggested by such moral norms as the principle of reciprocity and concern for children and the aged. However, the existence of such moral universals is ambiguous evidence: it may also be interpreted as a requirement for the viability of any human society. We shall regard it as evidence for both this thesis and the thesis that morals favor survival.

Yet there are plenty of cases that would seem to refute the hypothesis of the existence of moral universals: namely bellicosity, cannibalism, infanticide, and the neglect of young girls and of the aged. However, some anthropologists (e.g. Harris 1979) have offered a plausible explanation for some of these aberrations, as well as for a number of seemingly irrational taboos: they would originate in food scarcity. In times of plenty such practices are discontinued unless enforced by a priestly caste. And infanticide and child neglect have all but disapeared wherever family planning has become widespread.

What is the biological function of moral norms? They help shape social behavior favorable to the survival and, if possible, the well-being of the higher animals in a social group. This is particularly true of parental and filial affection, the respect of the rights of others (especially the right to life), and mutual help. From an evolutionary viewpoint the function of a behavior rule explains its adoption: The animals that fail to comply with the norm succumb early in life and leave no descendants.

Does this entail that biology is in a position to explain the origin and function of morality? Darwin's earliest champion denied it emphatically: "the ethical progress of society depends, not on imitating the cosmic process, still less in running away from it, but in combating it" (Huxley 1893 p. 92). On the other hand a modern-day Darwinian thinks that "the time has come for ethics to be removed temporarily from the hands of the philosophers and biologicized" (Wilson 1975 p. 562). Wilson and his fellow sociobiologists believe, in particular, that altruism, though possibly harmful to the help giver, is advantageous to its kin, whence it must eventually have been "encoded" in certain gene(s). (See e.g. Axelrod & Hamilton 1981, Alexander 1987.)

The following objections must be raised against this speculation. Firstly, it mistakes an inborn disposition for its actualization through experience (e.g. learning). Secondly, neither altruism genes nor hardwired neural networks for altruism have been identified. Thirdly, evolutionary biology is doing well without the kin selection hypothesis. Fourthly, developmental psychologists know that altruism can be stimulated or inhibited by rewarding or punishing feelings of empathy and helping actions — in short, it can be learned. And the fact that altruistic behavior is more likely to benefit relatives than genetic strangers can be explained simply by the fact that, for geographical and historical reasons, animals are more strongly bonded to their relatives than to total strangers (Panksepp 1986).

In conclusion, biology does not entail morality. If it did there would be a single moral code, whereas in fact human history is littered with moral norms fallen into disuse — e.g. many concerning honor and sex. What biology can do is not to crank out a definitive morality but (*a*) to identify the objective basic biological needs of human beings, (*b*) to evaluate scientifically certain customs (e.g. rites, taboos and deviations); (*c*) to join in the criticism of the ethical views that ignore or postpone basic human needs; and (*d*) to help criticize the extravagant claims that morality and immorality are inborn and, in particular, that certain races or social classes are morally superior to others.

3.2 *Psychological*

We are all familiar with moral emotions, such as compassion, generosity, trust, shame, and remorse. The scientific study of moral emotions, initiated by Darwin (1871), is nowadays a task of only a handful of social developmental psychologists — too important a problem to attract many students of the human mind. (See e.g. Piaget 1932, Goslan Ed. 1969, Bandura 1977, Aronfred 1980, Gilligan 1982, Kohlberg 1984, Moessinger 1989.)

Regrettably moral psychology has yet to make a deep inroad into moral philosophy. Many philosophers are still under the spell of Hobbes's false dogma that man is basically selfish and violent, though smart enough to realize that it is in his own best interest to strike deals that will favor his security and well-being. This pessimistic view of human nature, which ultimately derives from the Old Testament, was opposed to no avail by Spinoza (1677), Bishop Butler (1726), and a few others. It remained so influential that, when Adam Smith turned from moral philosophy to economics, he replaced his former view (1759), that sympathy is the foremost bond of human society, with Hobbes's opinion that the main spring of human action is egoism (1776). If anything, this false and socially dissolvent opinion is even more popular today, at least among philosophers, economists and political scientists, than when Smith advocated it. We shall attack it for several reasons, not the least of them being that people do have moral feelings, some of which motivate them to perform disinterested actions.

Darwin (1871), the pioneer of moral psychology, conceived of the "moral sense" as the feeling of right and wrong. He regarded it as a social instinct and as such inherited rather than acquired. This was heresy at a time when most moral philosophers believed that the "moral law" had a divine origin, or was imposed by the state, or was the product of a wily calculation of costs and benefits, or else derived from contracts struck by free agents. The anarchist Kropotkin (1902, 1924) made much of Darwin's view on the biological root of morality, insisting that "the first moral teacher of man was Nature". Likewise the socialist Kautsky (1906) wrote that the moral law is an animal impulse requisite for any kind of society.

A few examples should suffice to establish the existence of moral feelings and their role in the formation and implementation of moral norms. One is that in the world of children, regardless of their upbringing, "there is nothing so finely perceived and so finely felt, as injustice" (Dickens 1861 p. 92). It may be wondered whether contractarians, Kantians and libertarians have ever paused to think of the spontaneous emergence of the sense of fairness at a tender age, with little or no indoctrination — though certainly along with social intercourse.

Second example: it has been established experimentally that altruism is intrinsically rewarding — in particular, that people find it rewarding to deliver another human being from suffering (Weiss, Buchanan, Altstatt & Lombardo 1971). Third: When children between four and twelve years of age are asked to distribute a certain amount of goods among themselves, they invariably choose an egalitarian partition even if their "initial endowments" are different; the relevance of the latter, i.e. the notion of private property, only plays a role in some children above age thirteen (Moessinger 1989).

Fourth, there seems to be a natural (though not overpowering) inhibition of incest, not only among humans (Harvey & Tall 1986). Thus if given the option a mouse will choose to mate with a female that is genetically dissimilar to itself (V. Denenberg, personal communication). Psychologists have found that people prefer to consort with people that are different — though not too dissimilar. In particular, they choose friends and mates who are somewhat similar physically, psychologically and socially: they like difference and novelty, but within bounds. The incest taboos, marriage laws, and match-making rules seem to be rooted in the natural propensity to choose a middle way between close similarity and utter dissimilarity. The moral rules concerning incest would then seem to have deep biological roots.

In short, there are examples of scientifically studied behavior patterns guided by natural or spontaneous psychological inclinations. In particular, normal people, and perhaps other animals as well, have moral feelings that contribute to steering their social behavior. However, there is no consensus on what these might be. We attempt to capture this elusive idea by means of

DEFINITION 4.7 A moral feeling is an emotion elicited by the perception, memory, or imagination of an animal (possibly oneself), which one believes has need of help, or which one believes may cause or has in fact caused harm.

Some emotions, such as fear, rage, hatred, resentment, contempt, shame, and other "gut feelings", are processes occurring in the limbic system, which is mostly "hard-wired", i.e. the connectivity of which seems to be essentially determined by the genome: see e.g. Bunge & Ardila 1987. (However, the affect system can be somewhat modulated by cognition. For instance, one can learn to feel shame and contempt, as well as to control fear, rage, and hatred.) Other emotions, such as empathy, remorse, and respect, are largely learned (and sometimes unlearned). I.e., they emerge (or submerge) as a result of radical changes in neural connectivities in the course of individual development. (The control center of learned emotions is likely to lie normally in the left hemisphere.) Therefore such emotions must "implicate" not only the limbic system but also the cerebral cortex. We shall assume then that moral feelings are (*a*) processes occurring in the cortico-limbic system, (*b*) capable of being learned or at least honed or blunted, and (*c*) of social origin, i.e. caused by, or causing, social interactions.

Like all neural systems, the cortico-limbic one involves neurotransmitters that, by varying in concentration, alter the connectivity of such systems, i.e. the strength of their intercellular bridges. Other connectivity alterations are of an anatomical nature: they consist in the sprouting or pruning of synaptic boutons and dendrites. (Still, these morphological changes do not bridge the intercellular clefts: they only alter the intensity of the chemical connections.) Since any such changes modify our way of feeling, perceiving, thinking, and behaving, we shall have to look into them to discover the ultimate sources of moral feeling, reasoning, and acting. However, since brains do not work in a social vacuum, but are stimulated and inhibited by social inputs, it is likely that important alterations in social behavior, such as those caused by bereavement, change or loss of employment, and a quick increase or decrease in social interactions, affect the interneuronal connectivity, to the point that one may gain or lose in moral sensitivity or even in one's stock of moral norms. In particular, harsh punishment and the intake of psychoactive drugs are likely to cause a significant rise in the threshold of moral sensibility. In short, moral feelings are controlled from below

— the chemical level — as well as from above — the social level. Much the same holds for moral consciousness. See Figure 4.2.

Moral emotions can be studied physiologically at two levels: by tampering with the cortico-limbic system, or by recording peripheral electrophysiological responses. For example, one could establish whether or not a given moral emotion, such as empathy, readiness to help, envy, or shame, is localized in a given neural center and, if so, how it is affected by certain drugs. To this end one might temporarily inactivate the suspect center(s) by lowering their temperature to about 20°C, or one would inject certain drugs in them, while presenting the subject with a situation where another person, or an animal, is in distress. The same subject before the intervention could serve as a control. Such experiments would readily establish, among other things, whether the moral sense is lateralized. And they would suggest looking for them in homologous brain regions in other mammals, particularly other primates. The author has not come across any such studies.

The second kind of experimental study has been conducted on empathy, or feeling for others. The existence of empathic emotion has been established by combining self-reports with such electrophysiological measurements as that of skin resistance response. These studies have found, among other things, that (*a*) helping behavior is often motivated by empathy, and (*b*) imagining the target's feelings may increase empathy but is insufficient, by itself, to increase helping: knowledge of the other's distress must be "hot", not "cold", to move us to act (Batson & Coke 1983). Result (*a*) refutes Hobbes's thesis of the basic selfishness of human nature. Result (*b*) suggests that the formal teaching of ethics is no substitute for repeated exposure to human

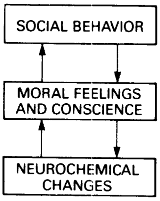


Fig. 4.2. Moral feelings, conjectured to be processes in the cortico-limbic system, are conditioned by chemical processes as well as by social circumstances. A radical alteration in either may elicit important moral changes. In turn these are bound to alter the subject's social behavior, in particular his motivation to perform right or wrong actions.

misery, let alone for active involvement in community or relief work. Presumably the individualistic and the negative utilitarians, as well as the contractarians, have been spared such experiences. For, if they had had them, they would know that many people are moved spontaneously to help people or animals in distress just because they empathize with them, and without pausing to make a cost-benefit calculation of the possible consequences, for themselves, of their actions.

Let us now move from moral feelings to moral consciousness, or conscience for short. (For the concept of consciousness see Vol. 4, Ch. 4, Sect. 5.1, or Bunge & Ardila 1987 Ch. 11.) We may define it as follows:

DEFINITION 4.8 An animal b has *conscience* (or moral consciousness) during the period $T =_{df} b$ thinks of its own moral feelings or thoughts (about moral problems, rights or duties) during T.

Like moral feelings, conscience is the prerogative of a few higher social animals. Machines can have a conscience only by proxy, and social groups never: there is no such thing as a collective conscience. The acquisition and alteration of a moral consciousness is part of the concurrent processes of mental (emotional and cognitive) development and socialization (enculturation). It begins by internalizing the moral rules prevailing in the group, and it expands by grappling with moral problems. (On the other hand according to Freud (1913) conscience would result from the confrontation of two inmates of the immaterial mind: the id and the superego. The least that must be said about this speculation is that it is untestable. The most is that, along with other psychoanalytic conjectures, it leads to obsession with oneself, hence unconcern for others, and therefore surrender of moral responsibility.)

Conscience may be modeled as a three-tiered system: moral feelings, moral reasonings, and internalized moral norms. The whole is constantly subject to variable social pressure. Moral feelings, though educable, are automatically evoked by external stimuli; in turn, they trigger moral reasonings controlled by moral norms. Moral reasonings stimulate some moral feelings (e.g. fairness and solidarity) while dampening others (e.g. envy and revenge). The three together, though in different proportions in different circumstances, trigger certain actions. The outcome of these actions may intensify or inhibit feelings, and reinforce or weaken norms, depending on whether or not the outcome is judged to be successful. See Figure 4.3.

Conscience becomes mature when the subject feels pleasure while doing the morally right thing, and guilty when he does not. When this

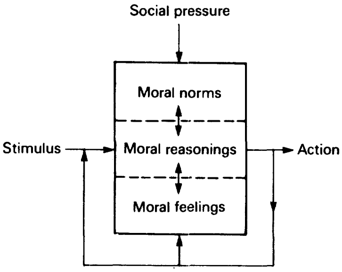


Fig. 4.3. Moral consciousness: composed of moral feelings, reasonings, and internalized norms subject to social pressure. The outcome of an action guided by moral feelings or reasonings alters the way the stimulus is "perceived" and it may hone or blunt the feelings as well as alter the norms.

stage is attained, moral conflicts tend to be felt as conflicts between prospective pleasure and prospective discomfort or even pain. Presumably, the pleasures and pains of the moral kind are processes localized in certain brain centers. It is also likely that moral conscience consists in the monitoring, by one neural center, of the moral feelings or thoughts going on in other centers: See Figure 4.4. However, conscience is not only a dashboard indicating the internal state of some parts of the brain: it is also an agent influencing our moral feelings, deliberations, decisions, and actions. In fact, moral decisions, unless spontaneous, are the result of deliberation, however quick, and they are likely to motivate behavior.

Example. The decision to try and save a drowning child by diving into stormy waters.

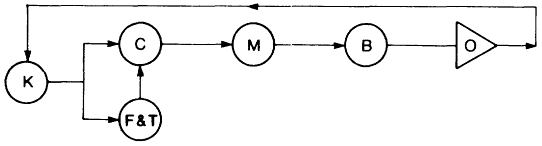


Fig. 4.4. Four neuronal systems involved in moral decisions and actions. Consciousness (C) monitors moral feelings and thoughts ($F\&T$) and both are partially controlled by knowledge (K). In turn, C is among the neural systems activating the motor systems (M) that control overt behavior (B). The outcome (O) of the latter is in part recorded by K. This feedback circuit allows one to update and correct one's fund of knowledge (including moral knowledge). Caution: This diagram is only a working hypothesis. Moral physiology is still only a wish.

Moral feelings, reasonings, and internalized norms are important because they motivate some social behavior: just think of helping behavior — or of aggression. Without conscience we would only be prepared to face routine problems, since these are the ones that promote the learning of automatic responses. Conscience and, in particular, self-consciousness, has the advantage that it helps us face totally new problems, for which our reflexes or even our experience are insufficient. In confronting unexpected situations we may either flee or take stock of our abilities, rights, and duties — i.e. to do what used to be called 'soul-searching'. Only such self-investigation can lead us to make rational moral decisions and to become better persons.

Without conscience it might still be possible to coexist with other people — but would it be worthwhile? Law and order can be kept by brute force or by calculated contract, whereas morals can only be enforced with a minimum of moral sense (conscience). If this sense is blunted, as it often is in the course of violent economic or political confrontation, particularly war, no atrocity comes as a shock, and no good action is appreciated: only winning matters. It is questionable whether people who have lost moral feelings, hence conscience, may be regarded as moral *agents* according to Definition 4.3. They are only moral *patients*. A dramatic example of loss of moral feelings and conscience in peace time, and one that made the headlines and has become a classic in social psychology, is the Queens murder. On March 13, 1964, thirty eight good people of Queens, N.Y., witnessed a woman being assaulted and repeatedly stabbed in the course of a half hour till she died. None of the witnesses bothered to alert the police, much less to intervene. They behaved the way individualistic utilitarians ("What's in it for me?") and contractarians ("What do I have to do with that?") are supposed to behave.

Fortunately most people do have a conscience, and most societies do have moral codes in addition to legal ones. However, moral sociology deserves a separate section.

3.3 *Social*

Moral codes are supposed to help spot and solve moral problems, which are subjective or interpersonal conflicts between rights and duties. Although some rights and some duties are biological and others psychological, moral norms are rules of social behavior: the hermit has little use for them. As such, moral codes are man-made, though seldom as consciously as tools or myths. Moral codes start inchoately and become increasingly explicit and definite as they are tried out and argued about. Eventually — sometimes once they have become obsolete — they get formulated and analyzed in detail, and perhaps even incorporated into legal codes. (More on the emergence and evolution of norms in Ch. 5, Sect. 3.)

Hence morality, regardless of its sources, is ultimately a social matter: It is supposed to regulate conflicts between self-interest and other-interest, between the individual and her fellow humans. In other words, moral codes are systems of norms of social behavior aiming at both self-protection and the preservation or improvement of the social order or structure. Hence no viable society, and no viable movement of social reform, can exist without one or more moral codes.

Moral biologism (Sect. 3.1) and moral psychologism (Sect. 3.2) overlook the social roots and functions of morality. The former ignores the fact that moral behavior is a special kind of *social* behavior, and that it is largely learned — so much so that different social groups are likely to practice and preach somewhat different moral codes. And moral psychologism ignores the fact that moral feelings — such as those of empathy and benevolence — and moral reasonings and sophisms, though necessary to spot moral problems and implement moral norms, are not enough. In fact any action motivated or hindered by the moral sense is done in some society and must count on means supplied by society (which is short for "other people").

Morals are born from the wish to satisfy the individual's basic needs and wants as well as those of her dependents. For instance, we all pay at least lip service to honesty because dishonesty breeds mistrust, which in turn weakens the social bonds; and some anthropologists (e.g. Murdock 1949) believe that most societies bar incest for allowing sexual competition within the family, thus causing its breakdown. Morality is then a device of social control as well as a guideline for personal behavior; another such device is the law. On the one hand the former is indirect: it acts through example and precept, through disapproval and the pangs of conscience. On the other hand the legal system acts directly by warning, deterring or punishing social behavior of certain kinds (ideally only those of the antisocial type).

Hobbes and Hume saw in morality what Mackie (1977) called the "invisible chain" that keeps us in our (deserved or undeserved) places in society, forcing us to obey and cooperate when we would rather rebel and defect. Consequently every breach of a moral norm, and every proposal to reform it, would amount to breaking or reshaping one or more links in the "invisible chain". This metaphor only sheds light on one side of the moral coin, namely duty. Actually every morality not only discourages actions of a certain type (mostly of the antisocial kind) but also encourages actions of other types (mainly prosocial ones). Morality is an invisible wing as much as an invisible chain: a winged chain rather than a chained wing. This could not be otherwise because humans, like the rest of primates, combine kindness with cruelty. As Wells (1916 p. 426) exclaimed in the middle of the Great War: "How kindly men are — up to the very instant of their cruelties!".

Every society has a single legal code (or rather family of legal codes), although it is often enforced differently according to the social status of the individual accused of breaking it. (For example, it is well known that, in racist societies, the members of the oppressed ethnic groups are usually punished far more severely than their oppressors if only because the latter can afford to pay for better lawyers.) On the other hand a complex society, particularly a modern industrialized one, has an entire family of moral codes: one for businessmen and another for workers, one for bureaucrats and another for the law enforcement officers, one for mathematicians and another for plumbers, one for priests and another for their faithful, and so on. These various codes are not always mutually compatible. The more complex a society, the larger the number of moral codes prevailing in it; and the more divided the society, the less compatible the various codes are with one another.

Sociologists have found that urbanization favors the tolerance of ideas, interests or practices that one does not share (e.g. Wilson 1985). There would seem to be two causes for this. One is the heterogeneity of the urban milieu, which would be unlivable without a modicum of tolerance. Another is the higher level of education of urbanites, who as a rule are more subject to school and media influences than small town people. Typically the latter are bigoted and more easily drawn by fanatic religious and political groups.

Like symbols, conventions, and machines, moral codes are neither natural nor God-given but human-made. However, moral codes are not artifacts of the same kind as time-pieces or computers, for, like languages, they are neither designed nor optional: they are social products and they are not easily altered, much less discarded. Still, they are human-made and so there is nothing sacred about them: they ought to be altered to suit new needs or legitimate wants. As Engels and Durkheim said, every society has the morality it needs — though sometimes social progress outstrips morals.

Yet, although morality grows out of social relations, nay is part of the social cement, it is not an epiphenomenon but reacts on social behavior, now constraining, now encouraging it. In other words, moral norms both "reflect" society and its changes, and contribute to the preservation or the reform of society. When the social group starts to decline without disbanding, mores and morals decline: old habits disappear, and so do solidarity, sharing, kindness, and tolerance. When morality dies, every family or even every individual fends for itself. The motto of the fascists *squadristi* was *Me ne frego* ("I don't care").

Having stressed the social root and function of morality, we must warn against moral sociologism. This is the thesis that morality is *nothing but* a social control device — a thesis first defended by Trasymachus (according to Plato in his *Republic*) and adopted in slightly different ways by Hobbes (1651), Montesquieu (1748), Engels (1878), Durkheim (1895), and Lenin and his followers (see Hörz & Wilke Eds. 1986). Moral sociologism is only partially true because (*a*) it ignores the biological roots of morality, i.e. the basic human needs; (*b*) it ignores the existence of moral feelings and conscience, hence the existence of inner moral conflicts; (*c*) it is incapable of explaining the moral differences between members of the same social class or occupational group — e.g. between liberal and conservative businessmen, or anticapitalist and procapitalist workmen; (*d*) it does not account for the moral constraints on business practices (e.g. those intended to protect the consumer); (*e*) it overlooks the moral constraints on political and military action (e.g. the "laws of war" included in the Geneva Convention); (*f*) by regarding ethics as *la physique des moeurs* (Durkheim 1895), moral sociologism mistakes morals for mores: it glosses over the differences between right and wrong customs, and (*g*) it underrates the power of moral criticism — which, after all, has often been a powerful social leaven, e.g. in the abolition of slavery and child labor.

The popularity of moral sociologism can be explained not only because it rightly emphasizes the social source and nature of morality, but also because it wrongly adopts the deontological or traditional view according to which morals are solely concerned with duties, authority, and respect. If, as in an authoritarian society, there are no rights alongside with duties, then there are no moral problems, hence no need for a moral code besides a legal one. A moral norm is such that its implementation is right and its breach is wrong — and the concepts of rightfulness and wrongfulness intersect but do not coincide with those of socially acceptable and unacceptable behavior respectively. Unlike natural and social laws, moral rules are optional, not binding. A person goes to jail only if he gets caught breaking a law while performing an action that may or may not conform to the prevailing moral code. For the righteous individual morals come before law and custom.

Finally, a word about morale, or the attitude of the members of a group toward the "cause" or goal they are supposed to be working or fighting for. A high morale is optimism together with dedication, whereas a low morale is pessimism (defeatism) together with slackness. Morale has an ambiguous relation to morals. A platoon out to kill the greatest number of "enemies", an organized crime group, or a gang of fanatics, may have a high morale accompanied by low moral standards regarding people not in the group. But, since maintaining a high morale requires mutual help, self-sacrifice, and discipline, a low exomorality is compatible with a high endomorality.

3.4 *Summary*

The function of moral norms is, to put it metaphorically, the protection of values. Since values are biological, psychological, or social, it should come as no surprise that morals have biological, psychological or social roots. Morals help solve moral problems, which are ultimately conflicts between the rights and the duties of a person, or between the rights or the duties of two or more persons. We would face no moral problems if we did not share the same niche with others, and if we were able to meet our needs and wants without the assistance of others.

Moral feelings and conscience play an important role in the recognition and solution of moral problems. This fact suffices to refute the purely rationalistic approach to morality, whether it be that of the contractarians from Hobbes (1651) to Gauthier (1986), or that of Kant (1785) and Kohlberg (1981).

Although moral norms have all three above-mentioned roots, they are rules of social behavior. They regulate the conflicts between self-interest and other-interest, i.e. the autonomy-reciprocity tension. If every individual were isolated from every other one, or if everyone were totally free to enter into contracts with others, or if our behavior were fully genetically predetermined, or if societies were organisms, there would be no moral problems, hence no need for moral norms.

CHAPTER 5

MORALITY CHANGES

Morals are not perennial but change along with individual and social development. They change because our needs and wants, hence some of our values alter as we grow up and as our societies become more (or less) urbanized, industrialized, cultured, bureaucratized, militarized, and so on. In particular, unlike our basic needs, our wants change rather quickly with the economy and the culture. Consequently our tertiary and quaternary values change as well, as a result of which our rights and duties change — changes which in turn alter our secondary norms. Recall Figure 4.1.

True, in most cases morals lag behind mores; i.e. moral changes do not follow automatically changes in their sources but come, if at all, somewhat later. In other words, moral inertia is greater than social inertia. Nevertheless moral changes do occur eventually in response to value changes, and sometimes as a result of technological or political innovation. For example, the popularization of the automobile has strengthened individualism, and that of oral contraceptives has shattered traditional sexual morals. And the threat of nuclear annihilation, accompanied by the quick deterioration of the environment, have awakened the social conscience of some people and dulled that of others.

In this chapter we shall take a quick look at the ontogeny and phylogeny of morals. That is, we shall examine the emergence of moral norms in the course of individual development from an early age on, and we shall study some of the changes in morals induced by alterations in lifestyle, e.g. from rural to urban.

We shall also study briefly the human condition. Are humans natural, artifactual, or both? And are we born good or evil? Next we shall study the two coins of human sociality: competition and cooperation. We shall argue that we need cooperation to survive, and competition to advance. The moral counterpart of this thesis is obvious: Enjoy life and help live.

1. DEVELOPMENT AND EVOLUTION

1.1 *Nature and Nurture*

According to a number of influential theologians and philosophers — among them Calvin, Hobbes and Kant — man is born evil and he can be made to behave decently only out of fear — of Hellfire or of the gallows, as the case my be. This moth-eaten view has recently been given a genetic twist: We would all be naturally selfish because organisms would be just machines serving the survival of genes (Dawkins 1976). But the selfish gene hypothesis has been rejected by most biologists because that which survives or succumbs is not the genome but the whole organism, and that which evolves is not the individual but the entire population or even the whole biosphere. (Recall Vol. 7, Ch. 3, Sect. 3.1.)

As for the hereditarian claim that we are born, not made, criminals or good citizens, it has no basis in either human genetics (Russell 1976) or psychology (Kamin 1974). The claim is based on a serious confusion between causation and correlation (e.g. criminality-race), and it is being used as an ideological weapon to justify racial discrimination, capital punishment, unemployment, and the dismantling of the relief state (Kamin 1986). The born-criminal thesis is itself criminal, and it makes a mockery of the ideas of criminal rehabilitation and of moral responsibility (since we cannot be responsible for our genomes).

Radical nativism died long ago in psychology and, in particular, in moral psychology, i.e. the psychological study of morals. (For the attempt to resurrect it in psycholinguistics see Vol. 7, Ch. 4, Sect. 3.1.) To be sure we inherit most of our genes, and these control, among other things, the development of the brain — the organ of emotion, cognition, moral behavior, and much more. However, the genome is only one of the two independent variables on which development depends: the other is the environment. Whatever we become is a result of the interplay of the two.

Barring certain serious genetic diseases, anyone can learn to become good or evil, nurse or warrior, teacher or thief, depending on the circumstances of his upbringing and the opportunities offered by his society. To be sure there are native propensities, but these have to be cultivated to become actualized — or neglected to prevent them from becoming realized. The cause of this is that a part of the higher vertebrate brain is plastic not rigid: We reorganize it ourselves as we feel and perceive, learn and act. (Recall Vol. 4, Ch. 4, or see Bunge 1980, or Bunge & Ardila 1987.)

No doubt, some people are more teachable (in particular conditionable), motivable and creative than others. But the point is that we are born educable not educated. In particular we are not born knowing any morals — nor, for that matter, anything but a few primitive motor responses to selected internal or external stimuli. From birth to youth, and from adulthood to old age, we develop morally as well as cognitively and in other respects. And this development, though made possible by inheritance, is steered by the environment. Aristotle (*Nichom. Ethics* Bk. II, Sect. 1) knew it: "none of the virtues arise in us by nature", but "moral virtue comes about as a result of habit". We are born with only the power to become temperate, courageous, just, etc.

Most of the serious students of moral development are enrolled in either of two rival camps: *cognitive-developmental* theory (e.g. Piaget 1932, Kohlberg 1984), and *social-learning* theory (e.g. Bandura 1977, Aronfred 1980). Roughly speaking, whereas the former lean towards Kantian rationalism, the latter are empiricist. (The ratio-empiricist synthesis is still in the offing.) The cognitive-development theorists claim that moral development is but an aspect of cognitive development: They focus on the way children speak about moral problems — not on the way they feel, and seldom on the way they act. They see moral development as an inevitable process punctuated by some leaps and, moreover, as progressive.

Piaget postulated two main stages of moral development, characterized respectively by egocentrism *cum* heteronomy, and by acentrism *cum* autonomy. The former is the *morale du devoir*, characterized by obedience; the latter is the *morale du bien*, characterized by mutual respect, fairness, reciprocity, and cooperation. As children develop they advance from the first to the second stage. (The theory, or rather hypothesis, fails to account for the cheats and criminals of the real world. Remember that Piaget's subjects were not New York street urchins but came from decent Swiss families.)

Kohlberg distinguished no less than six, and occasionally seven, stages of moral development. Every one of them would be inferior to the next — from the helpless yet spoiled brat to the fearless armchair saint. The first stage is that of blind obedience; in the second, the child observes the prevailing rules as long as they suit him; in the third he does what the others expect him to do; in the fourth he is law-abiding

— for what would happen if everybody were to break the law?; in the fifth he acquires a moral conscience: he distinguishes morals from mores, and learns that there are certain basic and universal rights and duties; in the sixth, which is optional and therefore rarely attained, the individual is a full-grown Kant who treats others as ends in themselves and adopts only universal moral principles. Thus moral psychology would justify certain moral maxims in addition to accounting for their emergence: It would have prescriptive as well as descriptive power.

The strongest point of cognitive-developmental theory is that psychological research has established beyond doubt the reality of *cognitive* developmental stages. Moreover there is neurophysiological evidence that the onset of a new stage consists in the reorganization or rewiring of neural circuits (Thatcher, Walker & Giudice 1987). The main trouble with the theory is that it overlooks affect, action, and social stimuli and inhibitors as well as social structure (Moessinger 1989). An additional problem with the hypothesis is that it underrates the role of moral inventions and conventions as well as the process of moral degradation.

C. Gilligan (1982) has criticized the cognitive-developmental theory for different reasons, namely for focusing on rights and justice while overlooking care and "relationships" (bonds). She regards this shortcoming as a product of male bias. The criticism may be correct but the explanation is not, for males too learn to care (particularly for their playmates) and they enter into "relationships" of many kinds. The neglect of care and bonds on the part of Piaget and Kohlberg had nothing to do with gender and much to do with Kantianism and with the fact that most of the available subjects seem to have been boys, as a consequence of which those investigators hardly addressed peculiarly feminine concerns such as those of infant care and male dominance. It is not that boys inevitably develop an "ethic of justice" and girls an "ethic of care". There is no morality at all without both fairness and concern for others.

Contrary to cognitive-developmental theory, social-learning theory claims that morality is purely a matter of social behavior: That we learn to observe some moral code or other in much the same way as we learn anything else, namely through conditioning, imitation, and precept. We "internalize" (make ours) a socially accepted norm of conduct when we learn to comply with it nearly automatically and feel that it is wrong for anyone to break it. Consequently moral development would be a gradual process steered by the social environment rather than by internal factors. Consequently there would be no qualitatively different stages of moral development.

The oldest argument for social-learning theory is found in Plato's *Theaetetus* 73, where Socrates says of a man that "His condition, which has been that of a slave from his youth upwards, has deprived him of growth and uprightness and independence; dangers and fears, which were too much for his truth and honesty, came upon him in early years, when the tenderness of youth was unequal to them, and he has been driven into crooked ways; from the first he has practiced deception and retaliation, and has become stunted and warped". (By the way, this stunning view contrasts with Socrates's cognitive nativism.)

The strongest argument for this view is that, after all, moral behavior is part of social behavior, which can only be learned in society. The acquisition of a moral conscience is part of the socialization process, and its deterioration is a concomitant of social degradation. The weak points of the theory, or rather hypothesis, are (*a*) the denial of developmental stages and the naive environmentalism peculiar to behaviorism, and (*b*) the inability to account for moral autonomy, in particular disconformism (especially acute in adolescence). The emphasis on extrinsic rewards ignores the experimental findings that (*a*) a person may retain her moral integrity even in the absence of reinforcements; (*b*) exclusive reliance on external incentives may undermine the internal ones (Lepper, Greene & Nisbett 1973); (*c*) making certain activities — such as helping, learning, and even playing — means to an end can likewise undermine internal incentives (Lepper, Sagotsky, Dafoe & Greene 1982); and (*d*) reasoning with the child and appealing to his sense of fairness and empathy is more effective than promising her material rewards or threatening punishment (Grousec & Dix 1986).

Clearly, then, each of the two main rivals in current moral psychology holds a grain of truth. In particular, whereas the cognitive-developmental school focuses on cross-cultural or universal features, the social-learning school attempts to account for the relativity of moral codes. Hence the two schools are in some regards mutually complementary rather than contradictory. Therefore it should be possible to unite them into a single theory. But such a synthesis ought to include three important components absent from the two rivals: (*a*) motivation and the moral feelings (such as empathy and shame) and reasonings (particularly those sparked off by the question 'What if everybody were to do the same?') that drive much of our moral behavior; (*b*) the neural "substrate" (organs) of such feelings and reasonings (and the moral deficits caused by their malfunction or destruction); and (*c*) the ideology accompanying every socially accepted moral code and every system of formal education.

The proposed synthesis can be summarized as follows. (i) Moral development is an aspect of biosocial development. (ii) The latter is nothing but the maturation of the entire body, in particular the central nervous system, in response to both endogenous (particularly genetic) and exogenous (mainly social) stimuli. (iii) Moral development is at the same time emotional (honing or blunting of moral sensibility), cognitive (internalization or criticism of moral norms), and behavioral (adoption or rejection of moral behavior patterns). (iv) Moral development does not proceed gradually but by qualitative leaps which correspond to massive reorganizations of the plastic regions of the brain. (v) Being a biosocial process, moral development can be steered (for better or worse) by the consistent application of stimuli of various kinds, in particular chemical and social — as suggested by the deleterious moral effects of drug addiction and the uplifting moral effects of model imitation. (vi) By the time she has reached adolescence, the normal human being has acquired a conscience that allows her to steer herself within bounds, and even to rebel occasionally against the prevailing moral conventions. (vii) Since conscience is consciousness of moral feelings and reasonings, the organ of morality is likely to be a comparatively large portion of the corticolimbic system. (viii) Because morals are learned they can be taught, particularly by example and by appealing to moral feelings, reasonings, and experiences. (See e.g. Chazan Ed. 1985.) And because the organ of morality is largely plastic, it is possible to reeducate moral deviants. (ix) The individual moral philosopher can do very little beyond his circle of relatives, friends and readers. On the other hand governments can lead millions of people to either moral uplifting or moral degradation, both directly through public education, and indirectly through the design and implementation of national projects engaging most of the population. A constructive project can inspire cooperation and selflessness, a destructive one will induce antisocial behavior. An honest government will teach honesty, a corrupt one corruption. (x) Given the overriding function of government in modern society, and the fact that every form of government is guided by some ideology or other, it is naive if not deceptive to overlook the ideological component of moral education both direct (or formal) and indirect (by example).

1.2 *Origin and Breakdown of Norms*

Since moral norms are contained in the collection of rules of social behavior, and since some of the latter are society-specific, it is reasonable to suppose that some norms differ from society to society and change over time in every society. Just think of the recent changes in sexual morality and in public attitudes towards the arms race and environmental pollution.

It is the task of the science of morals — in particular the psychology, sociology and history of morals — to identify the cross-cultural moral invariants as well as the intercultural moral differences; to trace the changes in morals, if possible from prehistoric times on; and to offer explanations of such differences and changes in terms of environmental, biological, psychological, and social (in particular economic, political and cultural) terms. Regrettably the science of morals, in particular the history of morals, is underdeveloped. To be sure there are plenty of studies on the history of ethical doctrines; but since these have been mostly pulpit or armchair cogitations, they tell us next to nothing about actual moral behavior in real society.

The problem of the origin, enforcement and withering of moral norms may be approached in a number of ways — preferably in several. First there is the *historical* approach, which attempts to trace the natural and social evolution of norms. A sensational result of this approach is the recent finding of fossil remains of hominids and primitive human beings who attained adulthood in spite of being severely crippled. This finding has been taken as circumstantial evidence for the hypothesis that our remote ancestors, far from being selfish and cruel — as Hobbes and his followers supposed — were capable of kindness. As for morals in historic times, there are plenty of pertinent documents. However, the value of such documents must not be overrated, because they begin to appear long after the norm in question has been adopted, and they may continue to bear witness to the existence of the norm even long after the norm is no longer observed.

The historical approach is most fruitful when combined with the anthropological and sociopsychological perspectives. For instance, the study of primitive societies and their transition to agricultural societies has led Opp (1982) to conjecture that the evolutionary emergence of norms goes through several stages, among them (*a*) the initiation of new behavior patterns that prove to be successful — or at least are believed to be convenient for some; (*b*) the explicit formulation of the corresponding normative statements; (*c*) the widespread acceptance and internalization of norms, and (*d*) the setting up of formal mechanisms for enforcing the norms. What begins as a spontaneous new behavior pattern, perhaps tried out by a single individual, may spread and become the (moral or legal) law of the land. Such combination of invention, trial, imitation, diffusion, convention, and sanction, is peculiar to moral and legal norms in contradistinction to natural laws.

Second there is the *observational* approach adopted by psychologists, anthropologists and social psychologists, who try to find out how norms emerge in the behavior of children, and how they are enforced or relaxed in primitive tribes, street corner gangs, and other social groups. This approach requires long follow-up studies in the case of individuals, and decades of observation of societies in transition. A defect of this approach is that the informants may be unreliable if only because, unknown to them, some norms may be in the process of emerging or of becoming extinct.

Third there is the *experimental* method: A social group is artificially set up by the experimenter, who assigns its members a task requiring interactions leading to moral conflicts. The major problem with this approach is that the groups are artificial and ephemeral, usually composed of individuals who are strangers to one another; moreover the groups are detached from society at large and they are usually assigned tasks that are quite different from those they confront in real life. Fortunately these are small hurdles in the case of children invited to play games, for usually they get together quickly, and the experimenter does not have to rely on their verbal reports. In fact direct observation suffices to find out, say, how the children divide loot and argue among themselves about the fairness or unfairness of the distribution. (See Moessinger 1975.)

Fourth, there is the *sociobiological* speculation, which attempts to trace the origin of norms to their biological value, genetic "encoding", and diffusion throughout a genetic pool. Such speculation might suggest how certain universal moral norms arise, but cannot possibly account for any norm bound to a particular place and time. Sociobiological speculation is fun because it produces the illusion of explaining everything, in particular egoism and altruism: the former for favoring individual survival, the latter for fostering kin survival. Besides being extremely accommodating, sociobiological speculation overlooks moral feelings and moral dogmas, and it does not account for moral criticism or even for plain conversation, which — unlike the shouting of orders — requires mutual respect and therefore a modicum of social equality. In short, whatever the merits of insect sociobiology, human sociobiology does not have many besides the good intention of supplying a scientific basis for social science and ethics. (More in Kitcher 1985.)

Fifth, there is the fashionable *game-theoretic* approach, often combined with sociobiological speculation concerning evolutionary "strategies". Its proponents seek to explain the adoption or breakdown of norms in terms of (positive or negative) payoffs in the manner of the Prisoner's Dilemma (single shot or iterated). An advantage of this approach is that the "norms game" can be simulated on a computer (Axelrod 1984). But this is also a shortcoming, for such simulation (*a*) ignores feelings and (*b*) it can only process the premises — in particular the payoff "data", actually invented — contained in the program: it cannot replace actual experiment or even observation.

Besides, a two-person game is an oversimplified model of any real social group, even of a childless married couple, for it overlooks the interpersonal bonds as well as the many relations that each individual has with third persons. Worse, it is centered on (short term or long term) self-interest, with total disregard for moral feelings, duties, and conventions. The players are supposed to be selfish and cool strategists — i.e. bastards rather than virtuous or even normal people. Finally, normative theories, such as game theory, cannot explain facts, such as the emergence or breakdown of moral codes, unless it be assumed that ours is the most rational of all possible worlds. To explain facts one needs theories including specific laws, and these are conspicuously absent from game-theoretic models.

Sixth and last, there is *psychoanalytic* myth making. In particular, Freud (1913) speculated that morality was originally indistinguishable from religion and sociality: All three would originate in infantile sexuality and, moreover, in the so-called father-complex and the associated fear of castration. Morality would boil down to the mastering of the Oedipus complex. Since this complex is imaginary, the psychoanalytic fantasy is just one more modern myth. (Recall Vol. 7, Ch. 3, Sect. 5.1.) Worse, it is an onerous myth, for it does not allow the clinical psychologist or the teacher to identify, let alone correct, moral deviations.

In sum, only half of the above mentioned approaches seem promising, namely the historical, observational, and experimental ones. They should be combined and practiced in the light of the hypothesis that morals have multiple roots, rather than pursued independently from one another or on the assumption that morality has a single source.

The practical utility of such studies should be obvious at a time like ours, when humankind is going through the most severe moral crisis in its history, when traditional values and virtues, whether good or bad, are being rejected without being replaced. This moral crisis is a result and in turn a minor cause of such deep social problems as overindustrialization, overpopulation, overurbanization, militarization, power abuse, and secularization. (It is not that morality is impossible without religion, but that the traditional moralities, most of them based on religious dogmas, have yet to be replaced with moralities compatible with contemporary science and technology.)

Since the current moral crisis is only one component of the total crisis that threatens the survival of humankind, only a systemic approach to that megaproblem could work. (Systems of problems have only systemic solutions.) This is not an excuse for postponing the building of a new morality up until the social problems have been solved. On the contrary, we need a new moral perspective to grapple with those problems. Only, this new morality, to be effective, cannot be the outcome of an academic exercise best left to moral philosophers working in sound-proof rooms. It must come from the combined effort of concerned scientists, technologists, scholars, educators, and men of action.

1.3 *Summary*

Moral norms are learned not inborn. They are learned along a complex course of physical (in particular cerebral) maturation and social intercourse. To understand this process one should study its various aspects at different levels. In particular, one should study its emotional, cognitive and behavioral aspects both in the individual and in her relations with her groups. Any unifactorial approach to moral development — such as the cognitive-developmental or the social-learning ones — is bound to produce some results while failing to understand the process in its full complexity. What is needed is an integrated or systemic approach to moral development.

Morality differs from one society to the next, and from one epoch to the next, even though all moralities proper share some fundamental principles corresponding to human nature. At least half a dozen different approaches to the study of the emergence and breakdown of moral norms have been tried. Only three of them — the historical, observational, and experimental — have produced tangible results, even though every one of them has its snares. The other approaches are far too speculative and simplistic. The severe moral crisis that humankind is currently going through cannot be understood, much less overcome, with the help of any such speculative and simplistic approaches. Being but one aspect of a global crisis, it calls for a multidisciplinary approach.

2. THE HUMAN CONDITION

2.1 *Human Nature*

Although no two human beings, not even "identical" twins, are strictly identical, we are all alike in most respects: This is why we all belong to the same biospecies. This platitude bears repetition in view of the persistence of extreme axiological and ethical sociologism. This view entails (*a*) radical *relativism*, or the denial that there are any universal human values and norms; and (*b*) axiological and moral *pessimism*, or the view that we cannot reasonably hope for valuational and moral reforms before society has been reformed. If there is such thing as human nature, however variable, then it must be possible to find cross-cultural values and norms, and to claim their priority in our quest for a good society.

All humans have a common ancestry, whence we have similar anatomical and physiological (in particular mental) features. Therefore we all have the same visceral and psychological needs — whence we all share the primary and secondary values. We all use and occasionally even manufacture artifacts. We all experience emotions, in particular moral feelings. Normally we all delight in learning and are naturally active. And we are all in need of help and normally enjoy company. All of these features together make us remarkably conflictive as well as cooperative beings.

Consider briefly two characteristically human tensions: those between passion and reason, and between selfishness and altruism. According to some (e.g. Hume and Pareto) reason is the slave of passion, whereas to others (e.g. Zeno the Stoic and Kant) reason must subdue passion. Each thesis is only partially true. In fact, we can often control emotion, and just as often use reason to satisfy passion. In particular, passion for a cause can fuel research and action, and analysis of feelings can check and even extinguish impulses. Developing into a mature person is in part learning to harmonize passion and reason: to use the former as a means for the latter and conversely. This learning is possible because of the strong anatomical links and physiological interactions between the cortical and limbic systems, and the interactions between these and the endocrine system.

As for the second conflict, our common needs make us all half-selfish and half-altruistic. As Dr Jekyll put it, "all human beings, as we meet them, are commingled out of good and evil" (Stevenson 1886). The evolutionary explanation of this mixture of good and evil is that those who were purely selfless (prosocial) did not take proper care of themselves and therefore did not live long enough to leave offspring; whereas those who were purely selfish (antisocial) either got no care from others when in need, or were eliminated by their antagonists. Shorter: To survive we must be minimally selfish and minimally altruistic. Even a few economists (e.g. Robinson 1962 p. 10) know this.

These and other conflicts, whether internal or inter-personal, make human life interesting, difficult, and tragic. Interesting and difficult because they pose tough problems, particularly of the moral kind. Tragic because some such problems are insoluble: Think of failure, estrangement, unrequited love, natural calamity, or death. Tragedy is part of the human condition. However, we can prevent much unnecessary tragedy, such as poverty, war, and all the miseries they cause. And this is what morals are or ought to be all about, namely the minimization of avoidable tragedy and the multiplication of opportunities for happiness.

Notwithstanding the common knowledge of the basic similarity of all human beings, hence of their basic needs and values, as well as of their basic conflicts, the very concept of human nature has been under heavy attack. It has been criticized on two major counts. One is that humankind has never stopped evolving, the other is that man is largely artifactual. However, the first criticism only refutes the simplistic thesis that human nature is constant: it does not establish the thesis that we have risen above nature. Of course human nature is not invariable — nor is elephantine or falconine nature. Still, we are animals belonging to the same biospecies.

The second criticism is more to the point for, far from being only the creatures of our genes and our natural environment, we are largely shaped by our social milieu — which we in turn alter through work. Man is the self-made artifact. In other words, the human condition is half-natural and half-artificial. Therefore the idea that man has at least originally been in a "state of nature" (as Hobbes and Rousseau believed) is false. So is the contemporary version of this thesis, namely human sociobiology.

Human sociobiology (Wilson 1978, Lumsden & Wilson 1981) is a variant of biological reductionism and, in particular, geneticism — i.e. the view that, in the main, human behavior is predetermined by the genome. Most social scientists have ignored or rejected this view for at least two reasons. Firstly, although no two persons are identical, almost any social role can be played by a number of individuals, to the point that in certain social groups, such as armies and big corporations, individuals are treated as expendable. Secondly, although all economic, political and cultural actions are instances of animal behavior, they differ across societies and from one historical period to another, without any appreciable accompanying genic changes. (Thus nobody would attribute the American Civil War to a macromutation — nor would hold that the event caused a massive genic change.) Genes make certain human actions possible and others impossible, but they are impotent without the assistance of the (natural and social) environment. (Recall Sect. 3.1.) Therefore, although ethics should take biology into account, it cannot be biologicized anymore than it can be psychologicized or sociologicized. (For further criticisms see Lewontin, Rose & Kamin 1984, and Kitcher 1985.) Finally, the sociobiological thesis that dominance and status are always positively correlated with reproductive success or fertility may be true of baboons and a few underdeveloped human societies but it is false of civilized societies. In fact the latter exhibit an almost inverse correlation between social (in particular economic and cultural) rank and fertility (e.g. Vining 1986).

In short, there is such thing as human nature because all humans, regardless of their individual differences, share a huge number of properties, among them those of neural and social plasticity. Were it not for such commonalities it would make no sense to speak of the unity of the human species. And were it not for the ability to change both mentally and in our social behavior, we would not be humans but hominids or even more primitive primates. Shorter: human nature is for real but it is eminently variable.

2.2 *Competition*

It would seem obvious that, self-preservation being the *summum bonum* of all animals, they are naturally selfish. And, given the scarcity of some resources, it would follow that they are naturally competitive and occasionally even aggressive. In particular, murder and theft, exploitation and oppression, war and conquest would come naturally to man — unless checked by weakness or by fear of retribution.

All this is in fact what Hobbes and his heirs, from the social Darwinists and Freud down to some ethologists (e.g. Lorenz 1970) and anthropologists (e.g. Tiger & Fox 1971, Chagnon 1988) have been saying about man. Nietzsche thought along similar lines and so did the fascist and nazi ideologists — who, incidentally, admired Nietzsche's unscrupulous Superman. Thus Mussolini (1932) wrote that the fascist repudiates pacifism and "looks on life as duty, ascent, conquest" — yet he despaired of ever converting the peaceful Italian peasant into a war hero.

It is worthwhile to recall Hobbes's social philosophy for, if accepted, it entails that morality is either pure cant or nothing but a clever device for social control, only more subtle than the jail or the firing squad. Hobbes's social philosophy can be condensed into the slogan "Everyone for himself and the State for all". This principle may be spelled out as follows. Firstly, we are naturally asocial or even antisocial: "men have no pleasure (but on the contrary a great deale of griefe) in keeping company, where there is no power able to overawe them all" (Hobbes 1651 p. 64). Secondly, men are naturally competitive and even aggressive: "every man is Enemy to every man" (ibid.). Thirdly, there is only one way to prevent the mutual destruction of individuals, namely by having them strike mutually convenient deals protected by an omnipotent State. Fourthly, the State, by imposing law and order, is convenient to all because it gives security; in particular it ensures the right to life and to private property. Fifthly, and consequently, we must support a strong (i.e. merciless) government that will guarantee the "social contract", which is the invisible umbrella protecting every person-to-person contract.

The social sciences have amply refuted Hobbes's pessimistic psychology and anthropology, thus seriously undermining his political and moral philosophies. Firstly, developmental and social psychologists know that human beings are sociable from birth, and that they acquire social skills as they develop. (One wonders whether Hobbes or any of his followers have ever raised any children.) Secondly, first a number of travelers and colonizers, and later on anthropologists, have refuted Hobbes's contention that "the condition of meer Nature" is that of "Warre of every man against every man". Most of the warlike tribesman have been found to live in internally peaceful communities, usually sharing equally whatever they can get hold of. All such primitive societies are stateless. And, though a few of them are warlike, most are not — particularly if they are not supplied with guns by the civilized. As Kropotkin (1902 p. 88) said, "Unbridled individualism is a modern growth, but it is not characteristic of primitive mankind".

Thirdly, social psychologists have found that aggression is not an automatic response to provocation: Whether or not a person responds aggressively to a provocation depends on whether or not he perceives that provocation as hostile. They have also found that aggression is far more frequent among socially uncouth and verbally unskilled persons than among others. Finally, we do not need much scientific research to establish that there are such things as sympathy, friendship, and love, as well as empathy, tolerance, and caregiving. A moral philosopher who ignores them not only distorts the truth about the objective grounds of morality: he also becomes a tool of exploiters, oppressors, and warmongers.

What holds for people holds, *mutatis mutandis*, for other animals: competition and aggression are limited to meeting basic needs. Thus biologists know that competition is often avoided by relocation, i.e. by moving to a new territory with similar resources. (In the case of humans relocation takes the forms of migration or job change.) Ethologists distinguish between territoriality (defense of the means of subsistence) and aggression. They also distinguish intraspecific from interspecific aggression. Sexual competition is an instance of the former, whereas predation exemplifies the latter. But sexual competition is limited to brief periods and is never deadly; and predation ceases upon need satisfaction. (Even in times of drought predators and their potential prey are seen to share waterholes.)

All social animals are born with mechanisms that contain intraspecific aggression. For example dogs, unless trained to behave otherwise, do not fight to the death: usually the defeated animal signals submission. With regard to gregarious animals of the same species, we must distinguish intra-group from between-group aggression. The former is usually contained when existent at all, for otherwise it would cause group breakdown. Between-group intraspecific aggression seems to be practised only by human beings. Among other animals the rule is tolerance among conspecifics.

Ethologists know also that, whether or not an animal is aggressive, depends on its lifestyle. It will be aggressive only if it makes a living by capturing live prey — and even so just when hungry or to feed its young. For example, the Siamese fighting fish is a predator: when facing a provocative stimulus it is likely to adopt an aggressive or a defensive stand; on the other hand goldfish, being scavengers, are normally peaceful. (See Otis & Cerf 1963.) Finally, we all know that domestic animals, particularly dogs, can be trained to become either aggressive or friendly toward strangers.

In short, the claim that humans are basically antisocial is a myth. To be sure all animals are competitive in some regards and within bounds. And, unlike most other animals, human beings can oppress and murder. But at the same time humans *need* not resort to violence among themselves, for they can best meet their needs and wants through work and mutual help guided by reason. (In particular, war is not in our genes and nowadays only fanatics and weapons manufacturers benefit from it.) This is not to say that it will ever be possible to eliminate all human conflicts: competition is natural and, within bounds, healthy. The point is not to eliminate conflict but to circumscribe it and to resolve it through fair and intelligent negotiation rather than through violence.

2.3 *Cooperation*

Cooperation or mutual help is as much a fact of animal life as is competition. This fact could not be seen by the early evolutionary biologists or by their contemporaries in the social sciences, who only saw struggle everywhere. A great merit of Kropotkin's (1902) was to accumulate evidence for, and to spread, the hypothesis that mutual help was just as common as competition in the animal world. (His mistake was to believe that, whereas competition can lead either to progress or to regress, mutual help leads automatically to progress. Actually either can lead to either: the outcome depends largely on the circumstances and, in the case of the higher animals, on the goals as well.) Mutualistic behavior, also called 'reciprocal altruism', is no longer a romantic dream: it is generally admitted by zoologists, ethologists and sociobiologists. (See e.g. Maynard Smith Ed. 1982 Ch. 6.)

Some biologists, in particular Hamilton (1964), have argued that mutualistic behavior is in the genes because natural selection favored those individuals that maximized the survival chances of their kin even at the expense of their own welfare. However, geneticists have yet to identify the putative mutualistic gene(s), and the hypothesis of kin or group selection is still *sub judice.* As for the game-theoretic models, according to which cooperation is an effective strategy in the "game" of life (e.g. Axelrod 1984), they are unconvincing because they make no use of any biological laws. (Recall Sect. 3.2.)

Whichever the final verdict on the hypothesis of kin selection may be for non-human animals, the evidence regarding humans is ambiguous. On the one hand it is true that kinship bonds are very strong, particularly in traditional societies. On the other hand sociologists have found that most acts of violence in industrialized societies, such as wife beating, child abuse, rape, and even murder, are committed against relatives or acquaintances. We also know that the level of violence varies from one society to another. (Compare Miami with Stockholm.) In conclusion, the hypothesis that prosocial behavior "is in the genes" is yet to be confirmed with regard to human beings. What we do know is that we are all born sociable, and that both prosocial and antisocial behavior can be reinforced or discouraged.

We also know that normal human beings (*a*) like to play and work together, loners being exceptional and not well looked on by the majority, and (*b*) like to help others and need help from others. Mutual help is particularly important in learning, especially in learning moral norms: We all learn from one another (elders and peers), and hope that what one of us does not know another may find out — this being why we keep asking for information and advice.

Competition alone would not have brought us to the present level of cultural development, for the learning process is essentially cooperative: We learn not only by ourselves but also from one another. It is most likely that the human brain coevolved with society, for the more complex types of social behavior, particularly team work and defense, demand a superior intelligence; and in turn the cultivation of intelligence leads to inventing or perfecting new and more efficient forms of cooperation.

We submit then that neither competition nor cooperation leads by itself to progress, whether biological or social: the former may lead to destruction, the latter to stagnation. We submit that social (in particular moral) progress is best achieved by *combining competition with co-*

operation. To be efficient, each must moderate the other: both violent competition and selfless cooperation threaten the survival of the individual as well as the integrity of the group. A major function of a good moral code should therefore be to restrain competition and stimulate cooperation. See Figure 5.1. We shall come back to this matter in Ch. 10, Sect. 2.4 and Ch. 11, Sect. 4.

Given that the components of any successful social group cooperate in some respects with one another even while they compete in others, it is distressing to find that social scientists are making increasing use of game theory, in particular of the Prisoner's Dilemma. This is distressing for the following reasons. Firstly, the case of two prisoners held incommunicado is hardly a realistic microsociety: in real life information flows are constituents of the mortar that holds society together. Secondly, the assumption that each of the two prisoners behaves egoistically ("rationally") overlooks their previous commitments, their moral feelings and moral codes, their fear of revenge, and their hope of resuming cooperation after leaving prison. Thirdly, since the entries (utilities) in the payoff matrices are assigned arbitrarily by the game theorist, he can "prove" anything he wants; in fact, whereas some game theorists have "proved" that defection (betrayal) pays, others have "proved" that cooperation does. (For more on utility theory and its applications see Ch. 3, Sect. 3.2.)

Even more distressing is the revival of the doctrine that "Violence [. . .] may be the principal driving force in the evolution of culture"

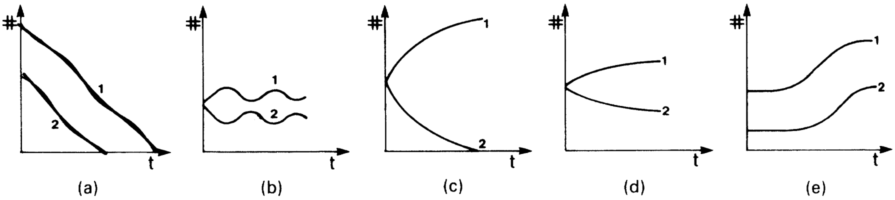


Fig. 5.1. Predation, competition and cooperation in nature and in society. (*a*) Unchecked rivalry: The two antagonists fight to the death. (*b*) Moderate predation: The numbers of prey (1) and predator (2) oscillate out of phase. (*c*) Unchecked competition for the same ecological niche: group 1 survives and overpopulates at the expense of 2, which becomes extinct. (*d*) Checked competition without cooperation: 1 and 2 coexist but 1 fares better than 2. (*e*) Cooperation combined with competition: 1 and 2 coexist and prosper by checking and helping one another.

(Chagnon 1988 p. 985). It is understandable that an anthropologist specialized in studying a bellicose tribe such as the Yanomamö, a sociologist expert in class warfare, an economist focusing on business rivalry, or a military historian, should form such a one-sided and bleak impression of the human group. But this impression is clearly false, for without a modicum of cooperation there would be no social groups. It is also dangerous, for the belief in the inevitability of violent confrontation facilitates conflicts of this kind and weakens the will to avoid or resolve them through negotiation or arbitration.

2.4 *Summary*

There is no question but that all animals are endowed with self-preservation mechanisms, some inherited, others learned. However, self-interest is not always conducive to antisocial behavior: in many circumstances it leads to cooperation. If an animal needs help in order to stay alive or raise its young, it will engage in mutual help, as in nest building and defense.

Cooperation is particularly conspicuous among social animals. In fact without cooperation — albeit mostly unwitting — there would be no societies at all. This is especially the case of human societies, whether they are primitive and surrounded by a hostile natural environment, or advanced and characterized by an extreme division of labor.

Human beings are at the same time competitive and cooperative. To give free rein to competition is to invite social disintegration, and to demand pure altruism is to court stagnation. The best moral and social policy is to combine competition with cooperation, i.e. to observe Norm 4.6 in Ch. 4, Sect. 1.2: *Enjoy life and help live.*

3. MORALITY CHANGES

3.1 *Evolution of Mores and Morals*

The human condition is not constant. Neuronal plasticity makes for mental, behavioral and social plasticity. In particular, values and norms change along with society, especially along with the cultural subsystem, of which values and norms are a part. Some norms are altered, others replaced; some are reinforced, others relaxed. Such changes may come about either because the values or norms have become obsolete, or because they are "perceived" as hurting the interests of some social group or, finally, because the society has entered a period of decadence and breakdown.

In civilized societies before the American Revolution (1776) and the French Revolution (1789) there were duties and privileges — sometimes conquered, at other times purchased — but there were no universal rights. However, in the *ancien régime*, at least in theory, those enjoying the most privileges and means had the most obligations toward the underprivileged, particularly widows and orphans: *Noblesse oblige*. But they enjoyed privileges which we now regard as outrageous, such as the *jus primae noctis* and the right to live in luxury without working.

In a political democracy there are rights and duties but, at least in theory, no privileges except for the very young, the very old, and the handicapped. Still, rights are largely confined to the political sphere and they did not come overnight and without struggles. For example, voting started as a privilege, it only became a universal right in the course of our century, and is now both a right and a duty (sometimes a legal one) in all democracies. Private property too began as a privilege, possibly about five millennia ago, and eventually it became a right, though one that is limited by law and in fact restricted to a minority of the world population. On the other hand procreation, probably considered as a duty (sanctioned by fertility rites) when the human population was scarce, is now a right. But, like the right to own land, that of procreation will soon have to be drastically curtailed all over the planet to ensure the survival of humankind: See Ch. 11, Sect. 2.

Any drastic change in lifestyle, particularly if it occurs across an important social group, is bound to be accompanied, followed or preceded by profound moral changes. For example, the recruit to industry from rural communities, particularly from tribal ones, comes with a moral code that "may contain no rules governing relations between employer and employed, buyer and seller, or workman and his mates [. . .] The clash of moral codes is painful" and it only adds to the difficult problems of adjustment to industrial life (Lewis 1955 pp. 195—196).

On the positive side we find, for example, the sudden availability of safe and comparatively inexpensive contraceptives from the mid 1960s on, which put an end to much misery, fear and hypocrisy — though at the same time it encouraged promiscuousness. An even more dramatic case was the refinement in moral sensibility which helped abolish slavery, serfdom and child labor. It would be naive to attribute such social and moral progress to economic causes alone, for the exploitation of slaves, serfs and children was very profitable. Likewise, the more humane treatment of criminals was economically less advantageous than their quick disposal. We must then admit that occasionally there is such a thing as moral progress, and that intellectuals and artists can do something to promote it — or to block it.

However, moral progress is neither automatic nor continuous: witness the horrors of world wars, extermination camps, and labor camps. Some contemporary societies have advanced morally and socially in some respects, e.g. in the treatment of ethnic minorities, women and children, while at the same time sinking into aggressiveness and corruption. This proves that there are no *laws* of moral evolution anymore than there are historical laws. In particular we are not bound to go either uphill or downhill in the field of morality. In each generation we become what we have chosen to become, and our choices are determined by our backgrounds, aspirations, means, and opportunities.

Our children and students become to some extent what we stimulate and allow them to become. On the one hand, if the accent at home, school and workplace is on obedience, discipline, and duty, a large mass of conformists and, by reaction, a handful of rebels will result. If on the other hand instant gratification, rights without duties, and unrestricted individualism are encouraged, antisocial individuals will result. If we wish moral progress we must learn to steer a middle course between authoritarianism (sheer dont's) and permissiveness (sheer may's), between the grim teachings of Calvin and Kant and the voluptous *Playboy* philosophy of life. We must learn to link rights to duties, enjoyment to discipline, creativity to discipline. We must learn to breed *us* generations instead of either *me* or *thou* generations.

To achieve this goal we must remember that moral education, like the learning of a skill, is both formal and informal: by example and exploration, by precept and argument, through understanding and emotion. Children, from about two years old, can be taught certain moral rules, particularly norms concerning fairness and mutual help. Somewhat older children can be engaged in argument about right and wrong actions, for they all have a keen sense of justice and they can readily empathize with their peers. Another important tool of moral education is play. A few games, such as playing house, are cooperative, but most are competitive: their "object" (goal) is to win. Still, whereas some of the latter (e.g. cops and robbers) can be aggressive, others (e.g.

tennis) are not. Still others, such as soccer, combine team work with inter-team competition and are subject to rules of fair play, whence they should be promoted. However, these and other sports become distorted and morally corrupt when commercialized: for the sake of public health and morals all sports should remain amateur.

In conclusion, rights and duties, as well as the accompanying moral norms, evolve with the societies where they are invented and which adopt or discard them. In other words, the evolution of morals, like that of the law, is an aspect of social evolution. This platitude should not be construed as a declaration of radical moral and legal relativism, according to which no moral or legal code is better than any other. There can be little doubt that some moral norms, e.g. "My country right or wrong", and some legal norms, e.g. "Theft is punishable under all circumstances", are bad and are being increasingly recognized as such. Nor is there room to doubt, except among the most backward-looking, that the *Déclaration des droits de l'homme et du citoyen* of 1789, and even more so the United Nations's Universal Declaration of Human Rights of 1948, have been moral and social achievements. So, moral progress, though not automatic or continuous, is possible. Hence it makes sense to ask for more of it, in particular to work for the world-wide respect for the 1948 charter of human rights as well as for the adoption of a Universal Declaration of Human Duties centered on mutual help both among individuals and among nations.

3.2 *The Current Moral Crisis*

Since the end of World War II the industrialized world has experienced two parallel but mutually opposed moral currents: one progressive (prosocial), the other regressive (antisocial). As mentioned in the preceding Section, a more honest attitude toward sex and a greater concern for the underdog are part of the former. But there are also clear signs of moral decline, such as a more ferocious individualism and a decline of the family, which used to be the main source of moral upbringing and support.

The most formidable obstacles to sustained moral progress are poverty and war. It is unrealistic as well as unfair to ask the jobless, the homeless or the hungry to observe any moral or legal norm other than that which recommends the mutual help among people in the same boat. And it is hard to expect anything but cynicism from people embroiled in a war that they have not initiated, or who live under the nuclear umbrella. Both the destitute and those who expect their lives to terminate at any moment are likely to feel that life has no "meaning" (purpose). No wonder, for they have not been consulted or given the opportunity to choose their own lifestyles. The only effective way to counteract such moral cynicism and despair is to rebuild society in such a manner that everyone may enjoy life.

Intellectuals, in particular moral and social philosophers, have had their share both in moral progress and in moral decline. Those who have written or campaigned for more enlightened and responsible moral attitudes, or for more advanced social structures, have contributed to the former. On the other hand those who have washed their hands of substantive moral issues, claiming that philosophers can only analyze moral utterings, but should abstain from moralizing, have been silent accomplices of moral decline. Finally, those who have exaggerated moral relativism or moral individualism, as well as those who have exaggerated moral absolutism or moral collectivism, have contributed to our current moral plight.

The two greatest philosophical enemies of moral progress are moral individualism — in particular nihilism and hedonism — and holism — in particular the view that all rights are vested in the state and all duties in the individual. Individualism involves the thesis of full personal autonomy and it makes people egoists like cats. Holism involves the thesis of full personal heteronomy and it makes people submissive like dogs. (More in Ch. 7, Sect. 1.7.)

The main intellectual enemy of moral progress in the West is moral individualism, which has grown enormously in recent decades despite the concomitant decline in opportunities to engage in private enterprise. (Philosophical individualism is not the same as respect for the individual: it is the defense of rights without duties.) The rise of moral individualism is not just the belated triumph of the ideology of greed. It is also a result of such social changes as urbanization (and the concomitant loss of the sense of belonging in a community) and the expansion of political rights, along with an easier access to education. It is also a result of such technological innovations as the automobile, the telephone, radio, and television, which have decreased the need for face to face contacts.

Even the computer is contributing to moral decline, though not for facilitating individualism but, on the contrary, for promoting our submission to an omnipotent and invisible Leviathan. In fact, the computer is increasingly taking both the credit and the blame for whatever goes on in a social megasystem. Although, when linked together and made accessible to individuals, computers help us solve a number of cognitive problems, in many cases they have become barriers between individuals. Where formerly one could tap the door of a supervisor or manager, and bang the door of a government agency, nowadays one is stopped by a clerk, or even a screen, telling us curtly: *Computator dixit.* See Figure 5.2. Not only this, but because computers have no moral feelings they cannot be compassionate and charitable: they are moral morons. As well, unless the computer programs are frequently updated, they become bastions of conservatism. In particular, some of the computer programs utilized in government agencies may come up with bad decisions based on instructions involving morally objectionable principles, such as "In matters of taxation always favor the rich", "In matters of employment always favor party members", or "In case of doubt shoot".

What can be done to favor moral progress? Preaching, whether from the pulpit or from the chair, is hardly effective these days. The only efficient moralizing actions are setting a personal example and actively participating in community endeavors (non-governmental non-profit organizations) as well as in politics. In modern society a moral person is not just a harmless nobody but someone who works for good causes together with similarly motivated people.

3.3 *Summary*

Morals change along with societies. They do not change automatically

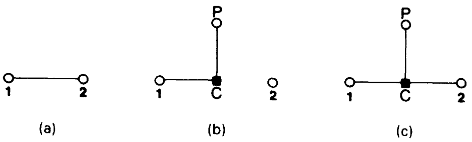


Fig. 5.2. Moral responsibility and the computer. (*a*) Yesterday: face to face interaction between persons 1 and 2. (*b*) Today: the computer *C* as a screen between persons and as a conveyor of instructions given by a third person, the programmer *P* (in turn a tool of the decision maker). (*c*) Tomorrow: individuals interact with one another and with the programmer through the computer.

or always for the better. Rather, they change along with lifestyles, which nowadays are largely shaped by industrial innovation.

The second half of the 20th century has seen two mutually opposite currents of moral change, one progressive, the other regressive. The former is obvious in the movements for human rights and environmental protection. Moral decline is obvious in the rise of individualism and corruption, as well as in the decline of the family and community ties.

Moral activism is inefficient when limited to preaching, but powerful when inserted in community and political activism.

CHAPTER 6

SOME MORAL ISSUES

In this chapter we grapple with some key contemporary moral issues. We do so in the light of the principles and definitions formulated in the two preceding chapters. In fact, this chapter may be regarded as a collection of examples aiming at elucidating and testing those postulates, norms, and conventions. However, as with all applications, we shall find it necessary to add a few assumptions and definitions.

Every selection involves an exclusion. In choosing some of the most important moral problems of our time we have ignored or minimized others which, though more hotly debated, are not central in our view. This is the case with abortion, animal rights, drug addiction, pornography, the treatment of the insane, street crime, and many others. To be sure these *are* real serious problems, but they pale by comparison with those raised by the arms race, political expansionism, racial discrimination, social injustice, environmental pollution, and underdevelopment. The importance of a moral problem is not measured by the publicity given it by the communications media or even by the number of scholarly publications devoted to it: instead, it is measured by the number of people involved in it.

Consider the following figures taken at random from a standard almanac (Sivard 1987). 800 million people are unable to meet minimal needs; 1,300 million do not have safe water to drink; 880 million adults are illiterate; the infant mortality rate in Sierra Leone is 198 as against 6 in Scandinavia; the life expectancy of an Afghan is 37 years whereas that of a Scandinavian is 77; the poorest fifth of the world population gets 1.6% of the total GNP while the richest fifth gets 74.2%; the U.S.A. and the U.S.S.R. together spend $1,500 million a day in "defense"; and the developing countries spend four times as much on arms as on health care. Add unemployment and underemployment, the continued growth of an already excessive population, the rapid depletion of nonrenewable resources, the dumping of toxic industrial wastes, acid rain, and particularly the threat of nuclear winter (or at least nuclear fall), and you have sketched the bleakest social and moral landscape ever contrived by man — in fact the worst of all possible worlds, to reverse

Leibniz's formula. Why moral as well as social? Because it is a result of personal greed — the lust for economic or political power — and because it will not improve unless we alter our policies and our individual behavior to conform with the principle *Enjoy life and help live*.

Our focusing on the most important moral issues of our time does not involve neglecting private morality: far from it. Public morality is not divorced from private morality if only because the moral agent is the same in both cases, namely the individual human being, and because one can hardly expect consideration for the public interest and for posterity from mass murderers or tyrants, crooks or profiteers bent on maximizing their own utilities rather than on rendering service to their fellow human beings. Private virtue is necessary for public welfare, which in turn renders private sin minuscule or even unnecessary.

1. PRIVATE MORALS

1.1 *Three Moral Spheres*

We may distinguish three moral spheres: the private, the professional (or occupational), and the public (or civic) ones. The private or personal moral sphere is the one ruled by norms supposed to hold for everyone regardless of her role(s) in society. The set of these rules, when good, usually goes by the name 'common decency'. It takes no professional training or public spirit to abstain from murdering, stealing or cheating, or to support one's dependents or to do a good turn for a friend in need.

The professional moral sphere is the one ruled by norms peculiar to each profession, such as not causing harm (on balance) in the case of the physician, and not leaving leaky pipes in that of the plumber. Since any society contains several occupational groups, each with its own specific moral code, we find a multiplicity of morals. However, the moral codes of all honest professional groups have a non-empty intersection, namely the norm "Do your job as best you can and do not take advantage of the weakness (physical, economic or cultural) of the people who employ your services".

Finally, the public or civic moral sphere is the one embracing our actions touching on matters of public interest, i.e. involving rights and duties regarding public goods, the community as a whole, or humankind.

The norms of public morality are the ones we are supposed to abide by as citizens, civil servants, community activists, politicians, or citizens of the world. See Table 6.1.

On the one hand, individualists are bound to claim that there is no difference between the three moral spheres because every moral norm regulates the behavior of individuals. On the other hand, holists are likely to hold that professional and public morals hover above personal morals and are independent of the latter. Each of the two extremes holds a grain of truth: all three moral codes concern ultimately individual behavior, but the public interest, when genuine — i.e. when geared to the well-being of all — precedes the private interest.

Still, the professional and public moralities are specializations of

TABLE 6.1. A sample of human private and public rights and duties.

	Rights to	Duties to
Private	live	help relatives and friends
	love	be fair and honest
	earn a living	support dependents
	rest and leisure	protect the weak and disadvantaged
	learn and investigate	teach
	move around	do one's work well
	privacy	keep the peace and prevent violence
	own what one needs for personal use	allow others to work, rest, and enjoy leisure
	choose place of residence	adopt a moral code of self help and mutual help
	social security	
Public	free opinion and expression	respect and protect others's rights
	associate	abide by the just laws
	elect representatives	protect public property
	run for office	protect the environment
	have access to culture	support good causes
	take part in the management of his workplace	keep well informed about public issues
	refuse to engage in acts of violence	help improve society, through both criticism and constructive work
	be judged by one's peers and by fair tribunals	pay taxes to support public services
	public services	fight warmongering, exploitation, and obscurantism

personal morality, for they take into account the special nature of the social group in question. Thus, the commandment not to embezzle public funds is just a special case of the rule not to steal, and the norm not to wage war is an instance of the rule not to kill people.

Consequently we reject the thesis of the autonomy of each moral sphere, defended e.g. by Weber (1921) and Freedman (1978). We claim that this thesis is wrong because it can be used to justify any crimes committed by *raison d'état*, from execution to occupation to conquest to genocide. Though distinct, the three moral spheres have a nonvoid intersection including such moral principles as "Enjoy life and help live", and the Golden Rule in some version or another. Consequently no maxim of professional or public morality ought to conflict with any principle of private morality if contradiction and hypocrisy are to be avoided. Thus one must not, *qua* professional or public servant, do what one ought not to do in private life. The lawyer ought not to break the law any more than he is free to cheat on his spouse; and the politician ought not to order the murder of anyone, whether compatriot or foreigner, any more than he is allowed to murder his own children.

There are several reasons for asserting the interdependence of the three moral spheres. Firstly, we wish to avoid contradiction and its moral correlate, namely hypocrisy. Secondly, we do not want to condone any crimes committed under superior orders or in the name of some cause allegedly higher that the common cause of human survival. (The Nuremberg trials ought to have settled this matter.) Thirdly, if we admit three mutually independent moralities we must concoct a fourth one in order to resolve any conflicts arising between any two of the former. But presumably such fourth morality would contain precisely those principles which we have placed in the intersection of the three moral spheres. (See Garzon Valdés 1986a for a similar argument.) Fourthly, to paraphrase Russell (1954 p. 28), without civic and professional moralities communities break down; without personal morality their continued existence is worthless or worse.

In short, there are three distinct but intersecting moral spheres: the personal or private, the professional or occupational, and the public or civic ones. Neither individualism nor holism can articulate all three into a coherent whole without sacrificing either the public or the private interest. The systemic solution is to combine individual rights with professional and civil duties in such a way that performing the latter makes it possible for the former to be exercised. No social welfare exists without individual well-being or conversely.

1.2 *Life and Death*

Since only the living can enjoy life and help live, our supreme moral principle (Norm 4.6) entails, among others,

COROLLARY 6.1 (i) Take care of your own life as well as that of others. (ii) Never put your life or that of others unnecessarily at risk. (iii) Never deliberately injure yourself or others. (iv) Protect your own life, as well as that of others, if possible by bargaining, if not by retreating, if not by defense, if not by counter-attacking. (v) Never strike first, and when you do strike back in self-defense or other-defense try to avoid killing. (vi) Refuse to take part in any killings other than those unavoidable in repealing armed attacks on home or homeland.

The right to life is the most basic of all rights, and the duty to protect life is the most basic of all duties. The basic nature of this pair is not only a matter of instinct or of morality: it is a matter of logic, for only the living can exercise rights or perform duties. Nor is the association of right and duty a contingent matter: One's right to life can only be assured with the help of others. Your (my) right to life implies my (your) duty to protect it. One's duty to protect the lives of others is usually performed by doing small services, often with no greater reward than a smile, rather than by heroic actions.

The effective implementation of the right to life in a world community with more than 5,000 million people is no easy matter. To live, a person's basic needs must be met, and this requires that certain economic, political and cultural conditions be satisfied. In particular, if every person is to live a normal life span she must have access to adequate food, drinking water, shelter, clothes, company, education, work, and health care. More than half of the world population lacks at least one of these bare necessities; as a consequence they cannot lead an active working life, much less a reasonably happy one. Therefore, proclaiming the right to life without at the same time doing something about the social order, is engage in a rhetorical exercise. Every right, in particular the right to life, is linked to other rights as well as to the duties implied by them. (Recall Theorem 4.1.) Hence, in order to protect or observe any of them we must also protect the associated rights and observe the associated duties. In complex matters sectoral thinking is bound to fail: only systemic thinking can succeed in grappling with whole systems of problems.

Our supreme maxim enjoins us to enjoy life not just to survive. Now, it is hardly possible to enjoy life unless one is free to choose one's lifestyle and occupation rather than having them imposed upon us without regard for our likes and abilities. Hence Norm 4.6 entails

COROLLARY 6.2 Everyone is entitled to choose her own way of life as long as it is neither harmful nor burdensome to others.

Thus everyone ought to be free to become farmer or cosmologist, computer programmer or brain surgeon, manager or even philosopher. But of course one won't get the help required to realize this ambition unless one shows promise and happens to live in a society where such help is available. Any career other than that of the unskilled worker calls for a rather long and demanding training punctuated by agonizing competitions: the freedom promised by Corollary 6.2 involves the possibility of failure. There is more: Since in every society some occupational groups are overcrowded whereas others are underpopulated, some measure of career planning is necessary. But such planning should consist mainly in encouraging — by means of fellowships and good job prospects — people to enter the most needed professions. And we must also be prepared to treat people willing to do disagreeable or dangerous jobs better than people whose task it is to read and explain novels or plays.

Most moral philosophers have been only concerned with the political, cultural and property rights of white adult males. Nowadays the rights of nonwhites and of women are recognized almost universally, though often only *de jure* rather than *de facto.* As for the rights of children, they pose a peculiar problem. They have not always enjoyed rights, not even the right to life. Infanticide, neglect, and the story of Abraham's sacrifice strike us now as barbarous and immoral; so do the data about child labor in Victorian England and in the European enclaves in China. The trend towards greater children rights is encouraging, but in the affluent countries many parents fail to teach their children that they have duties as well as rights. A teenager who feels that he has no duties whatsoever has been allowed to become an antisocial individual. Performing certain duties without being paid, in particular sharing in the housework, from the moment muscular coordination is attained, is simply being part of the family and becoming a useful member of society. Permissiveness, i.e. rights without duties, breeds moral monsters. And the idea that there can be education without coercion is a myth. "The task is not to undertake a vain search for the coercion-free educational system, but to consider ways in which the inevitably coercive aspects of socialization can be justified" (Ackerman 1980 p. 163).

So much for the right to life of living persons. Let us now deal with the unborn. This issue is covered by Norm 4.7: *Seek the survival of humankind.* This entails, among others,

COROLLARY 6.3 (i) Procreate without overtaxing the carrying capacity of the planet. (ii) Avoid *the* nuclear war (for it would almost certainly wipe out all human beings). (iii) Stop overexploiting the non-renewable resources and intensify the search for nonconventional energy sources as well as for extraterrestrial mineral resources. (iv) Clean up the environment, avoid polluting it any further, protect the forests, and reclaim the deserts.

Let us comment briefly on clause (i), leaving the rest for Sect. 2. In view of the current overpopulation, and that today about two billion people are not adequately fed, sheltered, educated, or gainfully employed, clause (i) must be taken very seriously on a planetary scale. The right to parenthood must be severely restricted, and planned parenthood must become a duty if humankind is to survive. It is up to demographers, economists and educators to figure out what the reasonable populations and birth rates should be in the various regions of the world; and it is up to the statesmen to adopt the experts's advice and have it understood and implemented by their communities. In short, the world population has become so huge, and the resources so strained, that parenthood has ceased to be a human right.

In contemporary civilized societies the right to human life extends normally from birth to natural death. Whether human embryos have the right to life is of course a disputed question. The affirmative opinion is based on the religious dogma that embryos have an immaterial and immortal soul insufflated into them by the deity. Why this dogma should be regarded as the ground for opposing abortion is incomprehensible: if the soul is immaterial and immortal, then it cannot possibly be harmed by its being separated from the body. Besides, contemporary child psychology has established beyond doubt that human embryos, or even fetuses, have no mental functions: that the latter start to develop only after birth. Before birth we are certainly human beings but not human *persons.* (Recall the Definition 4.46 of a person in Vol. 4, Ch. 4.) In short, the theological objection to abortion does not hold water.

Abortion should be avoided for a different reason, namely because it may be physically and mentally traumatic to the woman who has it done. The only way to avoid abortion, short of celibacy, is of course to avoid unwanted pregnancies, i.e. to practice scientific birth control.

However, if impregnation were to occur by accident, abortion should be permitted on the principle that the whole, if good, is better than any of its parts (Postulate 3.2). Moreover it should be recommended if the child is not wanted, or tests show that it would be severely handicapped. The reason for this recommendation is that it is morally wrong to make three people — the unwanted child and his parents — unnecessarily miserable. In any event, given that every year about ten million babies are born malnourished, fourteen million children die of hunger-related causes, and an estimated fifty million children have no shelter whatsoever, the issue of abortion is comparatively a minute one blown out of proportion by religious bigots.

The right to life extends to all persons, even to the worst criminals. There are two reasons for this view, one moral and the other practical. The former is that killing is always wrong, particularly so in cold blood: Two wrongs do not make a right. The practical reason is that statistics show that the death penalty is not an effective deterrent. There is a further reason, that is both moral and prudential, namely the slippery slope argument. Once you hang criminals of type A, you find it easy to hang those of kind B, and so on, till finally you hang mere suspects or rivals — as Hitler and Stalin did. For either of these three reasons the death penalty has been abolished in nearly all civilized countries. Crime is not to be punished but prevented. And, should it occur even when its causes have been eliminated, the offender should be reeducated as well as forced to do some socially useful work.

Since the right to life is vested in individuals, not societies, everyone has the right to take his own life or to ask to be put to death: life is *his* to enjoy or terminate provided he harms nobody by doing either. In other words, suicide should be regarded as a strictly private matter as long as the suicide does not leave destitute dependents behind. Everyone is justified in terminating his own life if he finds it unbearable due to intense chronic pain, loneliness, a feeling of uselessness, or unredeemable remorse. But of course all of the would-be suicide's acquaintances have the duty to try and change his circumstances so as to make his life worth living. Volunteer workers in suicide prevention centers know that a common sense piece of advice or a single affectionate gesture may suffice to change the would-be suicide's mind.

Euthanasia, or mercy killing, differs from suicide in that it involves a second party. It is justifiable under the same circumstances that justify suicide: when there is no more life left to be enjoyed. However, it is morally and technically wrong to throw the burden of mercy killing on the shoulders of a single person, such as a physician or a nurse. The decision to perform euthanasia should be reached after a rational and compassionate deliberation involving the patient (unless he has suffered brain death), his nearest relatives or friends, and a trio of professionals, e.g. a physician, a psychologist, and a social worker.

Murder, abortion, suicide and euthanasia are hotly debated issues because of religious dogma and because anyone can easily "identify" (empathize) with the victim. But they are very minor problems by comparison with war, famine and pestilence, which affect a vastly larger number of people.

Finally, what is the relation between morals and the knowledge of our mortality? This knowledge affects different people, with different cultural backgrounds, in different ways. Some take mortality as a reason for being virtuous or pious, others as an excuse for debauchery. Long ago, the *memento mori* induced some people to spend more time learning the *ars moriendi* than practicing the *ars vivendi.* Others, in particular the Stoics and the Buddhists, conquered the fear of death by becoming gradually detached from everything that makes life worth living. On the other hand the contemporary secular humanists believe that, because life is short, it should be lived to the full, and that we must become passionately attached and loyal to some people and certain causes, so that we may not regret having lived in vain. Take care of your *curriculum vitae*, and your obituary will take care of itself.

1.3 *Virtue and Sin*

If talk of virtue and sin sounds old-fashioned it is because some moral philosophers avoid using the words for fear of being taken for theologians, whereas others have surrendered to hedonism or even to moral nihilism, the natural companions of individualism. But of course moral virtue and sin, both private and public, are what morality is all about, as will be seen below.

What is true is that every culture has its own catalogues of virtues and sins. On the one hand, for example, the traditional Christian virtues were faith, hope, charity, temperance, prudence, justice, and fortitude. On the other hand the virtues extolled in a modern capitalist society are initiative, skillfulness, industriousness, competitiveness, prudence, acquisitiveness, and smartness. In a better society competitiveness would presumably be replaced or supplemented by mutual help, acquisitiveness by justice, and smartness by rationality. And the traditional

Christian virtues would not be dropped but updated: religious faith would be turned into faith in man, and hope for an afterlife into hope for a better world, but the other five would remain.

The list of sins too has suffered some changes. The traditional Christian deadly sins were pride, covetousness, lust, envy, gluttony, anger, and sloth. Except for the last, these look very much like the traits of character admired by the proverbial tycoon. On the other hand a secular moral philosopher might approve of the Christian list with only minor changes. For example, he might disapprove of pride in belonging to a privileged group, but might approve of pride in useful work well done, and might allow for lust but only in moderation. And he might also sympathize with Gandhi's list of sins: wealth without work, pleasure without conscience, knowledge without character, commerce without morality, science without humanity, politics without principle, and worship without sacrifice — only, he would worship, or rather respect, nature, in particular man, instead of supernature.

If we are to take virtue and sin seriously we must start by defining our terms and proceed to justify our preferences. We begin with

DEFINITION 6.1 (i) A *virtue* (or *sin*) is a trait of character that disposes her owner to realize a positive (respectively negative) value. (ii) A *moral virtue* (or *moral sin*) is a trait of character that disposes her owner to behave in a morally right (respectively wrong) way. [Recall Definition 4.2.]

Perseverance and diligence are nonmoral virtues, whereas fickleness and sloppiness are nonmoral sins — as long as their commission does not harm others. On the other hand altruism is a moral virtue whereas egoism is a moral sin. Our definition embraces only character virtues and leaves out ability virtues, such as intelligence and creativity. (See Rescher 1969 for this distinction.) The reason for this exclusion is that, from an ethical viewpoint, ability virtues, such as intelligence and discipline, are only means to moral virtue.

We shall deal only with moral virtues and sins, starting with altruism and egoism. Both come of course in degrees. In particular, altruism goes from solicitude to sacrifice, from lending a helping hand to laying one's life for another. We shall define 'altruism' as *concern* for the welfare of others, and 'egoism' as *exclusive* concern for one's own welfare.

Every known moral code enjoins us to be altruistic to some extent or other. One reason for this is of course that everyone needs some help from others: without some altruism no newborn would survive and no social group would last. Moderate altruism is only enlightened self-interest, a much more powerful social engine than sacrificial devotion to the welfare of others. On the other hand egoism is biologically suicide, socially corrosive, and immoral.

Our sole axiom about virtue and sin is this:

POSTULATE 6.1 Altruism is the highest moral virtue, and egoism the lowest moral sin.

The justification for this postulate is that altruism embraces all of the moral virtues and egoism covers all of the moral sins one can think of: See Table 6.2.

When saying that altruism is the highest moral virtue we mean that all of the other virtues are implied by altruism. The same applies to

TABLE 6.2. Some of the moral virtues and sins contained in altruism and egoism respectively.

Altruism	*Egoism*
Benevolence	Malevolence
Compassion	Cruelty
Considerateness	Inconsiderateness
Courage in facing moral problems	Moral turpitude
Courage in holding one's beliefs	Cowardice in holding one's beliefs
Discretion	Indiscretion
Environmental concerns	Environmental unconcern
Fairness	Unfairness
Filial devotion	Filial neglect
Forgivingness	Vindictiveness
Generosity	Avarice
Good will	Ill will
Honesty	Dishonesty
Helpfulness	Unhelpfulness
Loyalty	Disloyalty
Peacefulness	Violence
Public spiritedness	Civic indifference
Reliability	Unreliability
Respect for others's rights	Disregard for others's rights
Self-restraint	Incontinence
Sincerity	Hypocrisy
Tolerance (except for sin)	Intolerance (except for sin)
Trust	Mistrust

egoism. For example, altruism implies discretion and loyalty but not conversely: some people are discrete and loyal out of fear. Likewise, egoism implies indifference to environmental issues and civic apathy, but not conversely: an individual may be indifferent to environmental pollution and to political matters out of ignorance.

(Two points of logic before going on to substantive matters. First, virtues and sins are not necessarily mutually contradictory. Thus "peacefulness" has a single opposite, namely "violence". On the other hand the negation of "compassion" is the complex predicate "either indifference to other people's plights or cruelty". Second, the negation of a predicate is defined implicitly, namely thus: $(\neg P)x =_{df} \neg Px$. Likewise the disjunction of two predicates: $(P \vee Q)x =_{df} \neg Px \vee Qx$. In particular, $(P \Rightarrow Q)x =_{df} (\neg P \vee Q)x =_{df} \neg Px \vee Qx$.)

According to the Stoics and to Kant virtues are absolute ends in themselves and they are their own reward. For example, Kant held that it is wrong to lie for the love of another even if, in so doing, we would protect him from a murderer. Constant (1797) argued against Kant that there are no duties without rights; to tell the truth is a duty but only with regard to those who have a right to the truth; the murderer has no such right, hence we should not tell him where his prey is hiding. We agree with Constant. We do not have to choose between the Kantian *Fiat virtus, pereat vita* and the Nietzschean *Fiat vita, pereat veritas*. We opt for *Fiat vita, virtus et veritas*.

In the light of our supreme moral norm *Enjoy life and help live*, morality is not the worship of absolute values in themselves, regardless of circumstances, but a system of behavior rules subordinated to that supreme norm. This entails that virtues and sins are to be taken *in context* not in isolation. Thus loyalty is a virtue provided it serves higher values. For example, by betraying the trust placed in him by the dictator Franco, King Juan Carlos I of Spain remained loyal to his people. And by denouncing World War I, Russell, Einstein, Nikolai and a handful of others disobeyed their power-hungry rulers but were loyal to mankind.

No moral alchemy was at work in these cases: moral virtues did not get transformed into sins or conversely. What does happen is that there are no virtues or sins in themselves, just as there are no values in themselves. Instead, there are actions that, in certain circumstamces, or in view of certain goals, are virtuous whereas others are sinful. For example, it is right to help the needy or weak but wrong to help the wicked; it is right to forgive the repentant sinner, wrong to forgive the unrepentant one; and it is right to tell the truth to those who may make good use of it, wrong to tell it to those who are likely to make evil use of it. In short, virtue and sin are contextual or situational, not absolute.

Like rights and duties, virtues tend to flock together; sins are parallel. Any morally right action is likely to require further actions of the same kind; morally wrong actions are parallel. For example, performing a fair action involves good will, honesty, respect for other people's rights, and perhaps compassion and courage as well. In turn, putting compassion to work requires good will, generosity, honesty, and perhaps sincerity and courage as well. Both are examples of what may be called *moral virtuous circles. Moral vicious circles* are similar. Familiar instances are the use of violence to cover up dishonesty, of hypocrisy to perpetuate privilege, and of the disregard of others's rights to advance one's own.

Virtues (and sins) coalesce into systems, and these into supersystems. Thus sincerity, honesty and fairness together make up moral integrity. But moral integrity is compatible with a good measure of insensitivity to human suffering, hence with unwillingness to get involved in any good cause. Hence moral integrity is insufficient to make a fairly virtuous person. The virtuous person is as altruistic as it is possible to be without engaging in heroic self-denial.

Sins do not come in isolation but in systems. Thus the protection of crime involves lying, and the refusal to admit one lie leads to telling another. Virtues too coalesce into systems but not spontaneously: We must ensure that they do not conflict with one another. Thus kindness often comes into conflict with either justice or truthfulness; freedom, with efficiency or cooperativeness; loyalty, with fairness or with concern for humankind — and so on. Virtue conflicts can only be resolved either by deferring to the highest value involved or by working out a compromise.

1.4 *Ten Virtues*

In this section we analyze a mixed bag of moral virtues, genuine or bogus, which will make their entrance in alphabetical order. First comes *charity*, a spontaneous or calculated, individual or institutional action intended to relieve someone else's suffering. Shocking though it may sound, charity has ceased to be a virtue in modern times because (*a*) it helps perpetuate inequality and injustice instead of abolishing them and (*b*) it is sometimes used for such goals as climbing the social ladder, embellishing a corporation's tarnished "public image" or helping it save on capital gains taxes (Vickrey 1975), or achieving political domination — neither of which is morally justifiable.

The case of foreign aid is of particular moral, economic and political interest. It is certainly humanitarian in times of crisis (e.g. famine, flood, or pestilence). But when habitual it does more harm than good because (*a*) the recipient loses the habit of working to solve his own problems, (*b*) free food ruins whatever agriculture there was in the receiving country to begin with, (*c*) people get into the habit of eating foodstuff (e.g. wheat) that does not grow as well as their traditional cultivars, such as millet, sorghum, and rice, and (*d*) dependence on foreign aid strengthens political dependence.

The solution to endemic food shortages lies in a combination of endogenously generated social (e.g. land) reform with foreign technical help aiming at improving native agriculture. Such technical help may consist in (*a*) helping improve methods of irrigation, fertilization, harvesting, seed selection, and desertification control; (*b*) helping set up farmers's cooperatives for the production and direct sale of native grain to the urban population; and (*c*) helping improve the educational system, not only by teaching new (but appropriate or soft) technologies, but also the virtues of self-reliance and mutual help. Clearly, no such agricultural development plan can be efficient unless designed and implemented in close cooperation with the local authorities. In short, the only efficient and genuinely altruistic kind of charity is the proverbial "Help others help themselves". This is the only kind of charity with dignity. (For technical specifics see e.g. Spurgeon Ed. 1979.)

What is dignity? We identify it with respect. More precisely, we stipulate

DEFINITION 6.2 A person *b*

(i) is *treated with dignity* $=_{df}$ *b* is treated as a human being with rights and duties rather than as a tool;

(ii) *behaves with dignity* $=_{df}$ *b* behaves with self-respect, i.e. does not relinquish any of her rights and duties out of greed, fear, or the wish to conform.

To behave with dignity is not the same as to behave in a "dignified" manner, i.e. to pretend that no harm or inconvenience is done. Nor is it to put up with pain or loss: this is stoicism.

Forgivingness comes next. We should forgive the repenting sinner and we should never take revenge — which not only the ancient Greeks and Romans but also the *Old Testament* (*Deuteronomy* 32—35) and St. Paul (Romans 12: 19) regarded as a divine pleasure. But it would be unrealistic to ask anyone not to hate his enemies, as Christ and Spinoza wished. This would be biologically impossible, for hatred is one of the oldest brain functions. What we can do and ought to do is (*a*) to keep hatred under rational control, and (*b*) to work for a society where rivalry is not accompanied by violence. Either calls for good will.

Good will is the disposition to do good. According to Kant (1785) good will is a good in itself, regardless of the consequences. Popular wisdom in this regard, encapsulated in the saying "The way to Hell is paved with good intentions", knows better. Beware of the fool's good will: although his intentions may be excellent, his actions may be disastrous for lack of skill or knowledge. Only the intention-means pair may be attributed a full value. The value of any one of the members of this couple may be cancelled by the disvalue of the other.

Helpfulness is of course the spontaneous disposition to help others attain their goals. When reciprocal and limited to socially admissible goals, helpfulness is called *solidarity* or *fraternity*. Together with participation, the sharing of basic values and beliefs, and communication, solidarity makes up the mortar that holds social groups together. As Tocqueville (1835 Part II, Ch. IV) remarked long ago, social stratification hinders solidarity, whereas equality favors it. Hence, if we value solidarity we should promote equality. This is one more example of the inseparability of morality and the social order.

Honesty has several components: intellectual (truthfulness), political (service and transparency), and economic (not stealing). Let us confine our attention to the latter. Theft poses at least two moral problems: Whether it is always morally reprehensible, and when it is, why it is. According to an absolutist morality, such as Kant's, theft is always abominable. Not so according to a situational morality, i.e. one that takes the agent's circumstances into account. According to this view theft is morally justified on the conditions (*a*) that the agent commits the offense only occasionally, and to survive or to help someone else survive, and (*b*) that the action does not jeopardize the victim's survival.

What holds for ordinary theft holds for the expropriation of land or some other means of production. Expropriation is morally justified when (*a*) it is necessary to improve the well-being, and even more so to ensure the survival, of a group of people willing and competent to work the concern in question, and (*b*) if the previous owner receives a compensation or is invited to join the new enterprise. The right to life takes precedence over property rights because ownership, though sufficient, is not necessary to meet any basic needs. (More in Ch. 11, Sect. 3.)

Our second moral problem concerning honesty is this: Why should I refrain from stealing except to survive? The lawyer answers: To stay out of jail. The economist: Because it is in your own best economic interest. The individual psychologist: Because you have been conditioned to behaving honestly during your socialization period, and feel pity for your would-be victim. The social psychologist: Because theft generates mistrust and fear from sharing. The moral philosopher: Because theft is unfair to the victim. Who is right? They all are: There are several reasons for behaving honestly in most cases, even though one of them may be overriding in a given case. In the case of public property there are further reasons: Selfish, such as the wish to continue to have access to it or to avoid having one's taxes raised, and altruistic, such as public spiritedness. Both are justified.

Humility used to be regarded as a virtue. Nowadays we tend to think that a humble person underrates herself to the point of losing her self-respect and of refraining from taking credit or initiative. Humility must not be mistaken for modesty, or realization of one's unavoidable limitations with regard to both competence and rights. Typically, the original scientist, technologist, or artist, is modest but not humble: on the contrary, he is ambitious. There is nothing wrong with the ambition to attain something good, everything wrong with total lack of ambition or with the ambition to aggrandize oneself at the expense of others. Here as elsewhere the right path lies between the extremes: self-restraint is the ticket.

Self-restraint lies between self-indulgence (or incontinence) and self-denial (or self-mortification). The self-restraint of the virtuous person derives from his modesty (or consciousness of his own limitations) or from his concern for others, who would be harmed if he were to indulge in excesses, particularly in what Veblen (1899) called 'conspicuous consumption'. We owe it to ourselves (for reasons of health) and to others not to produce or consume anything in excess, particularly in view of the increasing scarcity of natural resources. Frugality is one of the virtues involved in a survival morality. Consequently inciting to conspicuous consumption and tolerating waste and wanton destruction are serious sins. In general, too much of any good thing is bad.

Self-restraint ought to apply, in particular, to sex. Here again the right path lies between abstention and profligacy: between self-denial and self-indulgence. The rationale for this norm is again our supreme maxim *Enjoy life and help live*. The first conjunct invites us to enjoy sex, the second to procure pleasure rather than harm to our partner. A sexual practice is morally objectionable only if it causes harm (bodily, mental or social) or displeasure to at least one of the persons involved in it. Our principle suggests that sexual virtue is identical with good old original sin in any of its healthy and pleasurable varieties.

By definition, original sin involves monogamy. Polygamy and polyandry are unfair, hence immoral, in societies with normal sex ratios because they leave some people unmated. But where the sex ratio deviates considerably from the normal it might be prudent to allow for a temporary deviation from monogamy. (Historical examples: Germany after the Thirty Years' War and Paraguay after the war against the Triple Alliance.)

As for homosexuality, there is no doubt that it is abnormal, but not more unnatural than disease. The right moral and practical attitude of heterosexuals toward homosexuality is (*a*) to admit that it is a deviation from the norm caused by either hormonal imbalance (hence in principle corrigible with testosterone injections) or by wrong upbringing (hence in principle corrigible with behavior therapy); and (*b*) to tolerate it without encouraging it among adults, while trying to prevent or cure it among children in the same way we deal with polyomielities or with selfish habits. Incest is in a similar boat: (*a*) it is avoided by the great majority of mammals and birds, which seem to value diversity (Harvey & Talls 1986), and (*b*) it is socially corrosive because it destabilizes the family and, unlike exogamy, it blocks the formation of larger alliances.

Next comes *sincerity*, which is likely to be peculiar to man. Deception — hiding, mimetism, mimicking animals of different species or even plants — is a rather common practice in the struggle for animal life. It is likely that man recognized rather early on that sincerity is essential to group stability, for it generates trust, without which there can be no social cohesion. On the other hand deceiving members of certain groups is still often regarded as praiseworthy. Plato, Voltaire, Nietzsche and James advocated employing useful ("royal", "noble") lies to keep their fellow men at bay. This policy may pay in the short run but it is sinful and socially corrosive.

Contemporary career diplomats are taught never to lie — but never to tell the whole truth either. They practice tact — dissimulation, self-restraint, and tolerance. Tact greases the wheels of social intercourse. It must be distinguished from hypocrisy, which is the faking of feelings. Whereas tact is beneficial for reducing unnecessary friction, hypocrisy is harmful because it allows the agent to occupy a place that he does not deserve.

Both tact and hypocrisy must be distinguished from moral inconsistency, i.e. the doing of what one finds morally reprehensible, or the refraining from doing what one believes to be morally commendable. (For example, the present writer believes in the benefits of vegetarianism and of political participation, but he eats meat and he rarely does more than paying his party membership dues.) Because moral inconsistency is bad for one's conscience and generates uncertainty and mistrust in others, we should reform society so as to minimize the number of occasions for indulging in it.

Our last item is *work morals.* Whereas *Genesis* (III, 17) regarded work as "the painful expiation of sin", Luther and Calvin regarded it as virtuous. (For a history of the subject see Tilgher 1930.) Several reasons have been given for the modern attitude toward work. According to some, work is moral because it is the only honest method for getting rich or at least avoiding poverty. (False: Millions of people work hard without rising above the poverty level.) Others claim that work is good because it gives a person a chance to create. (False: As a matter of fact most work, even in science and in art, is drudgery.) Still others claim that work is good because it helps your neighbor. (Not always: Working in an arms factory, or in a junk food or junk art factory, harms your neighbor.) The truth is that only useful and harmless work is good for self and others, whereas voluntary habitual idleness is sinful because it is selfish.

Since work is good, nay indispensable for survival, every able-bodied person has the duty to do some work. But in order to discharge our work duty we must have the right to work. Regrettably, nowadays millions of people cannot exercise this right: they do not find jobs. In the advanced countries the state takes care of the unemployed by doling them out just enough money to subsist. Such compensation is the least evil, hence it should be taken as a stop-gap. We should aim for a society where the relief state is unnecessary because everyone can earn his up-keep and that of his or her dependents; this is so because unemployment causes moral disintegration and unemployment compensation is charity. The aim should then be neither to extend the welfare benefits (the liberal wish) nor to shrink them as much as possible (the conservative dream), but to render them unnecessary. A social order incapable of guaranteeing the right to work is at best inadequate, at worst immoral. (More in Ch. 10, Sect. 1.4.)

1.5 *Desert and Reward, Crime and Punishment*

The concepts of good and evil, and those of virtue and sin, are necessary but insufficient to judge whether a given action deserves reward, punishment, or neither. To this end we need, in addition, the concepts of erogatory, supererogatory and undererogatory action. Doing one's duty is erogatory; this is all that the laws and regulations can demand in fairness. Doing more than what is strictly required is called a 'supererogatory action'; and doing either less or the contrary of what is required, may be called an 'undererogatory action'.

It is generally accepted that doing one's duty does not call for any special reward. Only acting beyond the call of duty entitles the agent to a special reward. And failing to do it or worse — i.e. sinning actively and harmfully against someone's basic rights — deserves punishment. However, such deserts are subject to moral as well as social constraints: rewards should not consist in privileges paid for by the misery of others, and punishments should not be destructive. More precisely, we propose

NORM 6.1 (i) The exercise of basic rights, and the performance of basic duties, deserve no special rewards or punishments. (ii) Supererogatory virtuous actions deserve reward, and undererogatory sinful actions deserve punishment. (iii) A just reward of the virtuous meets one of his legitimate aspirations. (iv) A just punishment of the sinner does not put his life at risk and it aims at his reeducation and rehabilitation.

A consequence of the adoption of this principle is that in any society some people fail to realize some of their aspirations either because these are unrealistic or because they pass unrecognized. The principle is consistent with the cult of the self-made man and inconsistent with the "entitlement theory" prevalent in the times of slavery and serfdom, and recently resurrected by Nozick (1974). As well, it is incompatible with the doctrine of collective responsibility and its attendant collective rewards and punishments.

A social group wherein merit is not recognized and rewarded, is doomed to stagnation, and one wherein crime is not punished is bound to decline or even disintegrate. As the English saying has it, where everybody is somebody, nobody is anybody. In such a radically egalitarian (or mediocritarian) society the word 'important' becomes meaningless, and there is little incentive to to anything outstanding or to take on any initiative. For example, a community of scholars where proofreaders, translators, and editors are placed on the same level as investigators and writers is bound to discourage or even harass its creative members.

In a society where the principle *Enjoy life and help live* prevails, the most meritorious actions are those which save lives or enhance the quality of life, and the worst crimes are those that cause the loss of lives (either directly by the force of arms or indirectly through poverty) or depress the quality of life. In such a society the saints or heroes are those who devote their lives to constructive causes, whereas the villains are those who are dedicated to destruction, oppression, or exploitation.

Murder is the worst of all crimes because it denies the right to life, which is the most basic of all rights. This is why most moral codes condemn murder. However, few do so consistently: in fact many codes condone legal murder, i.e. the death penalty, and the massacres perpetrated in the name of an ideology or in the alleged interest of a nation. That is, they condemn retail private murder but condone state mass murder — which, to say the least, is contradictory.

Although every murder is reprehensible, it is the more serious, the greater the number of victims. Accordingly we may rank the seriousness of homicidal acts: single person murder, small social group (e.g. family) murder, large social group (e.g. ethnic minority or political party) murder, war, and omnicide — i.e. the murder of every single human being through nuclear war. On this scale political terrorism, whether committed by a band of fanatics or by a state, lies midway between manslaughter and war. Though morally repugnant and politically inefficient in the long run, political terrorism has been blown out of proportion by the very political rulers who have engaged in state terrorism. As a matter of fact political terrorism is peanuts in comparison with genocide and war, which are always state controlled.

The state has, or ought to have, the right and the duty to protect human rights — in the first the right to life — and to force the observance of basic duties. Regrettably there are some criminal states, i.e. states where "law and order" forces specialize in jailing or assassinating political opponents, labor organizers, or intellectual leaders —

or in preparing and waging aggressive wars. Hegel and the legal positivists notwithstanding, might does not make moral right. Unfortunately most moral and legal philosophers have failed to address these larger issues.

The ordinary courts of criminal law only handle retail murder. They lack the laws and the power required to arraign the politicians or the brass responsible for state terrorism, genocide, or war; and of course there would be nobody left to judge the culprits of omnicide. Only some grassroots organizations, in particular political parties, have or may gain the power necessary to prevent or stop large scale murder. (For example, the Vietnam war was stopped by a popular antiwar movement.) This suggests that Aristotle was right in holding that ethics must be allied to politics. (More in Chs. 11 and 12.)

1.6 *Summary*

We distinguish but do not separate three moral spheres: the private, the professional, and the public ones. Correspondingly we distinguish private from public rights and duties, though without detaching them. For example, concern for other people's rights should be a crucial concern in all three spheres.

The most basic of all rights is the right to life, and the most basic of all duties that of protecting life. These rights and duties are to be understood in a broad manner that includes the right to choose one's way of life as long as it is neither harmful nor burdensome to others.

The right to choose one's way of life entails the right to abort an unwanted child. However, given the traumatic effects of abortion it should be avoided in the only rational way, which is by birth control. The latter will have to be universalized in order to check the suicidal population explosion.

Every morality is about virtues and sins. A moral virtue is definable as a trait of character predisposing to morally right behavior. Moral sins are parallel. In the last analysis all moral virtues are one, namely altruism; and all moral sins are one, namely egoism. Virtues and sins are not absolute but contextual or situational. For example, it is right to forgive the hungry person who steals a loaf of bread.

Some dispositions that used to be regarded as virtues, such as charity and honesty, are limited. The former because it may perpetuate helplessness and social injustice, the latter because the hungry may on occasion be forced to steal. Every candidate to virtue or sin must be evaluated in the light of the supreme norm *Enjoy life and help live.*

Supererogatory deeds deserve reward, whereas crime deserves punishment. However, the virtuous should not be rewarded at the expense of others, and punishment should aim at rehabilitation rather than revenge.

Murder is the worst crime, and mass murder — in particular war — is the worst of murders. The effective prevention of murder calls for social reforms, not just penal legislation. This holds in particular for the prevention of mass murder. Ethics must eventually coalesce with political science, and morality with political action.

2. PUBLIC MORALS

2.1 *Human Survival Issues*

Until recently most moral philosophers, with the notable exception of Aristotle, Spinoza, and Russell, could pretend in good faith that they could and even should remain above the political fray. The threat of nuclear world war, the rapid degradation of the environment, and the equally fast depletion of non-renewable resources, should have put an end to political innocence.

The Rev. Theodore M. Hesburgh (1983) has stated rightly that the most important and pressing of all the moral problems facing us today is that of the nuclear arms race for, unless stopped, it may put an end to all human life and therefore to all moral problems. Something similar may be said about biological warfare. Next in urgency come the twin problems of environmental degradation and of depletion of mineral (in particular oil) resources, for without a clean environment we shall not be able to live healthily, or even at all, and without mineral resources our descendants will be forced to return to the Stone Age. For these reasons we posit

POSTULATE 6.2 The most serious and pressing moral and practical tasks of our age are (i) to destroy all the nuclear and biological weapons, and to effectively block any attempt to resume their production, testing, and use; (ii) to clean up the environment, stop pulluting, and dispose rationally (e.g. by recycling) of the industrial and domestic wastes; and (iii) to halt the unbridled exploitation of the natural resources — particularly of the non-renewable ones — and to accelerate the search for alternative resources.

There are, to be sure, plenty of other pressing moral and practical problems, such as those of poverty, political oppression, and cultural deprivation, which affect roughly four-fifths of the world population. However, they cannot possibly be addressed unless humankind survives. Besides, they might be resolved within the next generation if only the current military expenditures, incurred at the tune of $1.8 million a minute, were diverted to good causes. Last, but not least, the three world-wide problems referred to in Postulate 6.2 concern every single human being regardless of nationality, class, or ideology. The moral problems to be tackled in the subsequent sections do not disturb the sleep of the privileged and they shade into ideological issues.

2.2 *Equality*

Marked inequality generates unhappiness, low self-esteem, envy, greed, dishonesty, anomie, dissatisfaction with the social order, and their social manifestation: uncooperativeness, violence, and eventually rebellion and its bloody sequels. The more marginal the individual, the less bound he feels to observe the prevailing moral and legal codes. For example, where gypsies are discriminated against they feel free to steal from members of the ethnic majority, whom they rightly regard as aliens, although they would not dream of stealing from one another (Martínez-Selva 1981). In short, the downtrodden are socially and morally anomic, and a deeply divided society is potentially explosive.

In a stratified society, particularly one lacking social mobility, it is only natural to behave in an egoistic manner toward members of other groups, either to survive or to retain privilege. In such a society altruism, reciprocity and mutual help occur only among members of the same social group, i.e. among equals. On the other hand in an egalitarian society, where everyone has roughly the same rights and duties as anyone else — hence similar benefits and burdens — there is no justification for witholding cooperation, much less for rebellion. Hence if we wish social cohesion and the effective rule of a single moral code dominated by the norm "Enjoy life and help live", we had better strive for an egalitarian social order.

Equality is necessary for attaining further social desiderata as well. To begin with, liberty can only flourish among equals for, as soon as some individuals become significantly more powerful than others, they are bound to take advantage of the weaker members of society. (See Tocqueville 1835, Part II, Ch. I.) Fraternity is parallel: If we wish to be treated in a brotherly way we must be effectively brothers or sisters, not poor cousins or rich uncles. (See Tocqueville 1835 Part III Ch. IV.) So is justice: Global fairness, i.e. economic, political and cultural justice can only exist among equals. In a divided society much energy is invested either in redressing grievances or in power abuse. Finally, integral democracy is in the same boat for the same reason; as a matter of fact integral democracy is hardly distinguishable from equality. See Figure 6.1.

(On the other hand the preservation of world peace and of the environment are largely independent of the equality issue, for they are in the interest of everyone. Moreover they are far more urgent tasks than that of attaining equality — which, after all, can also be attained instantly through nuclear war or gradually through environmental degradation.)

The problem then is not whether we ought to seek equality, but what kind of equality is best conducive to enjoying life and helping people live. To begin with it is obvious that, the American Declaration of Independence notwithstanding, all humans are not born equal: we only end up by being equal. The point is not to pretend that we are all born equal, much less to attempt to force biological equality (e.g. by cloning a single individual), but to preserve biological variety while preventing natural inequalities from developing into social inequities. In short, biological egalitarianism is impossible; and, even if it were possible, it would be undesirable.

Nor is radical or literal social egalitarianism desirable. For one thing it would be unfair to the very young and the very old as well as to the handicapped and the sick, all of whom require extra care. For another it would discourage, or at least not encourage, any supererogatory actions. Those who render distinguished service to society, e.g. by performing dangerous or dirty jobs, deserve rewards beyond their

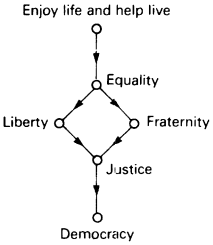


Fig. 6.1. Equality as a precondition of liberty and fraternity, which together make moral (social) justice possible, which is a prerequisite for (integral) democracy.

rights. What we want then is an equality that recognizes special needs and merits, as formulated in

NORM 6.2 (i) All human beings have the same rights to use the natural, economic and cultural resources of their society as they see fit and as long as they respect the same rights of others. (ii) All human beings have the duty to do their best to take care of themselves and to contribute to the well-being of their fellow humans as well as to pursue the survival and advancement of humankind. (iii) The sole inequalities justified in the distribution of goods and services are those which are to the benefit of all — namely the rewards of merit and punishments of misdeeds in accordance with Norm 6.1.

We may call this the *principle of qualified equality* or, in Ackerman's (1980) felicitous phrase, *undominated diversity.* It is a moral principle for concerning basic and universal rights and duties. Clause (i), the principle of resources equality, is a version of the norm of equal opportunity. It lies behind Kant's categorical imperative, or principle of the universalizability of all moral and legal norms. (In fact, Kant's principle cannot hold among unequals.) Put negatively, clause (i) enjoins us not to demand privileges, i.e. special rights without corresponding balancing duties. And the principle entails that no person may own any other person: it condemns slavery and serfdom, whence it justifies rebellion against slavery and serfdom.

Clauses (i) and (ii) together constitute a version of the socialist principle "To each according to his needs, from each according to his abilities". On the other hand clause (iii), which supplements Norm 6.1, is a modified version of a much criticized maxim due to Rawls (1971 p. 62). It is motivated by the obvious proposition that it is in everyone's interest that the most delicate, dangerous or dirty jobs be entrusted to those who are competent and willing to do them, and their holders be suitably compensated for their extra burdens — as long as they retain their abilities and be subject to control in order to avoid the abuses deriving from the concentration of power.

Clause (i) talks about *access* to, or *use* of, natural or social resources, not of private ownership. (Ownership implies right to use but not conversely.) Consequently clause (ii) cannot be interpreted as the doling out of gratuities to the less fortunate, and (iii) cannot be interpreted as unequal distribution of property rights. In fact (ii) talks about universal duties regardless of wealth, and (iii) about unequal distributions of jobs, in particular public offices, that not everyone is competent or willing to hold.

Private property is a legal not a moral category, for — *pace* Hospers (1982) — ownership is sufficient but not necessary to satisfy basic needs or legitimate aspirations. It has not always existed, and nowadays it does not exist in many social groups, such as primitive communities, cooperatives, and the state. Locke (1690) held that nobody has the right to own anything natural, because the earth is the common property of humankind; but that everyone has the right to "the fruit of his labours". Given the division of labor inherent in modern industry, and the fact that it uses natural resources — which belong by right to humankind — Locke's principle may be regarded as an *avant la lettre* justification of some sort of common ownership.

Wherever there is private property there is bound to be exploitation of some kind, from slavery to the self-exploitation of the self-employed farmer, craftsman, or trader. There is also likely to be violence in defense of one's property or in the attempt to expand it at the expense of others. (The success story of colonial empires, large corporations, and organized crime has a parallel though largely unwritten history of slavery, misery, death, ignorance, and other companions of the lack of property in a world ruled by the propertied.) Rivalry over property rights — on land, trade routes, and people — is likely to be the most powerful source of civil and international conflict.

What holds for private property in general holds in particular for the exclusive ownership rights to irreproducible goods, such as land and mineral deposits. Such rights are morally unjustifiable for a number of reasons, among them the following. (See e.g. Sandler G. 1977.) Firstly, as long as land is privately owned, the great majority of people are deprived from the space they need to live in good health. Secondly, landowners benefit effortlessly from the work of all those who live and toil around the former's properties. Thirdly, the real income of a worker is eroded by the high rent (usually one third of the salary) he must pay the landowner. Fourthly, unless property rights are curtailed by severe environmental protection regulations, the landowner is free to turn his piece of land into a toxic waste deposit that may poison air and water. Fifthly, the aesthetic enjoyment of nature is severely restricted. These five reasons suffice to indict the property rights on land. Land, particularly in an overcrowded world, should become the common property of humankind, only to be leased to firms, preferably cooperatives, as long as they make rational and socially beneficial use of it. (See Ch. 11, Sect. 3.)

So much for equality as a moral imperative. (For a vehement praise of inequality see Mussolini 1932 p. 49.) We shall not devote a special section to fraternity because it is implied by equality. In fact, fraternity is necessary for genuine equality; equivalently, equality implies treating others as brothers or sisters. Therefore the famous three word slogan of the French Revolution (1789) is logically reducible to *Liberté, égalité*. Only, in our view the correct order is *Egalité, liberté*, for liberty can only exist among equals. Which takes us to the threshold of the next section.

2.3 *Freedom*

We adopt the distinction between *passive* freedom, or freedom *from* constraints, and *active* freedom, or freedom *to* act (*potestas agendi*). The latter implies the former because an agent cannot act freely unless he is free from certain constraints, either because these are absent or because he has overcome or skirted them. Freedom, whether passive or active, can never be total, for every agent is under a number of external (natural and social) constraints. See Table 6.3.

We characterize active freedom, or liberty, by means of

DEFINITION 6.3 A person x is

(i) *free* to do $y =_{df} x$ is capable of doing y & no external constraint prevents x from doing y & x possesses or can obtain the means to do y & it is possible for x not to do y;

(ii) *morally free* to do $y =_{df} x$ is free to do y & neither y nor the means to do y are incompatible with the basic rights and duties of x or anyone else, or with the Norm 6.2 of qualified equality.

TABLE 6.3. A sample of desirable personal and social, passive and active freedoms.

	Personal	*Social*
Passive	from poverty	from discrimination
	from disease	from economic exploitation
	from fear	from political oppression
Active	to love & be loved	to form a family
	to learn	to join associations
	to work	to attend schools

In any society everyone is free to perform certain actions but not others: a modicum of freedom is necessary for survival, and some restrictions on freedom are required to live in society. However, in a society of unequals the powerful are freer than the rest. Only in a society of equals do all enjoy the same freedoms because everyone has access to all the resources of society (Norm 6.2). In such a society the only limits to a person's actions are the freedoms of others, and his own ability and effort.

Because any freedom is the ability to exercise a right or perform a duty, there are as many kinds of freedom as kinds of right or duty. Hence, in agreement with Table 4.1 (Ch. 4, Sect. 1.1) we distinguish five families of freedom: environmental, biopsychological, cultural, economic, and political. See Table 6.4.

A good society guarantees freedoms of all five kinds, not just one or two. Why? Because, like rights and duties, freedoms make up a *system*, i.e. the effective exercise of one renders the others possible and conversely. For example, civil liberties are needed in order to work and, if necessary, fight for the other four freedoms. In turn, without environmental, biopsychological, cultural, and economic freedoms, there is little point in political freedom. Politics is not a goal in itself but a means for the exercise of nonpolitical freedoms. Political freedom has done little for hundreds of million underfed, underemployed and undereducated Indians. Freedom does not nourish, employ, or educate, for it is mainly a means: recall Definition 6.3. The goal is equality in enjoying life and helping live, and equality is the only effective means for securing freedom, which in turn is required to safeguard equality. Thus, far from being mutually exclusive, equality and freedom support

TABLE 6.4. Five types of moral freedom and one example of each.

Type	*Example*
Environmental	To make fair use of the environment.
Biopsychological	To procure the well-being of self and others.
Cultural	To learn and teach.
Economic	To work for a salary.
Political	To participate in governing bodies.

one another. (See Dworkin 1978, Ackerman 1980, Nielsen 1984, Nino 1984.)

Freedom is necessary for self-actualization: "Man does not want liberty in order to maximize his utility, or that of the society of which he is a part. *He wants liberty to become the man be wants to become*" (Buchanan 1979 p. 112). (Caution: No one has the right to become an antisocial person.) But freedom is also an end in itself, for being free "feels good", as anyone knows who has been deprived from it. So, freedom is a goal as well as a means.

But of course *complete* freedom is unattainable and undesirable because (*a*) we cannot escape natural laws, (*b*) we cannot relax all of the natural and social constraints, and (*c*) we must observe certain norms if we wish to belong to any social group. In short, all freedom is limited. And a good society only admits *socially responsible liberty*, i.e. liberty to do right not wrong. As Russell (1949 p. 25) put it, "too little liberty brings stagnation, and too much brings chaos".

We have taken it for granted that freedom, however limited, is really possible, not just a theological or philosophical illusion. Our justification for doing so is the bankruptcy of the externalist dogma that the outputs of a concrete system are solely determined by its environmental inputs regardless of its internal structure and of the processes that occur in it. (See Vol. 4, Ch. 1, Sect. 3.1.) Limited self-determination or spontaneity is for real, as can be gathered from such diverse fields as mechanics (inertia), optics (self-propagation), chemistry (self-assembly of atoms or molecules), biology (spontaneous neuronal discharges), psychology (creativity), and social science (rebellion).

In short we reject the deterministic denial of free will (hence of freedom). However, we do not accept the indeterministic equation of freedom with lawlessness or arbitrariness either: on the contrary, we hold that free will (hence freedom) is compatible with lawfulness. Moreover we hold that the rational exercise of freedom calls for the existence and knowledge of laws: in a chaotic (lawless) universe there would be no room for forecast, hence for planning, hence for rational action aiming at the exercise of any personal or social freedom.

To sum up, freedom is possible at least to a certain extent. It is also desirable, for only free persons can effectively exercise their rights and perform their duties. Moreover only free human beings can be distinct individuals (rather than numerals in a concentration camp), fully responsible for their own actions and omissions, and capable of feeling self-respect. But genuine freedom is only possible among equals.

2.4 *Justice*

We distinguish two concepts of justice: substantive or moral, and formal or legal. The latter is adherence to the law, whereas moral justice is adherence to some moral code. The difference between the two is brought home on reflecting that the elaborate and logically admirable system of Roman (actually Byzantine) law was grossly unfair to the majority of the population for enshrining slavery, whereas most primitive societies are (socially) just yet lack in an explicit system of formal justice. Obviously, only a legal order matching a moral code is morally binding. To think otherwise is not to understand that there may be unjust laws, and that the cause of moral justice involves repealing or updating such laws.

In recent times a number of moral and legal philosophers, notably Rawls (1971), have equated justice with fairness though without defining the latter and often subordinating justice to freedom — as if there could be genuine freedom without equality. We propose to characterize fairness through

DEFINITION 6.4 A person x is being treated *fairly* $=_{df}$ x is being treated exclusively on the strenth of x's rights and merits.

Put negatively: A person is given an unfair deal if some of her rights are ignored or if she is discriminated against because of her physical appearance, sex, social status, or ideology. According to the above definition fairness is not an independent variable (or a primitive concept), for when judging whether a given person is given a fair deal we take into account her rights (hence her basic needs and legitimate wants) and her competence or ability (hence her deserts).

In a rights-only perspective it is natural to equate justice with fairness. In our rights-and-duties perspective fairness, though necessary, is not sufficient for justice. The owner of a plantation may treat his slaves fairly, but slavery is an unjust social condition. We postulate that justice involves equality as well as fairness, and disguise this postulate as

DEFINITION 6.5 A person x is being treated *justly* $=_{df}$ x is being treated fairly & x enjoys the benefits, and carries the burdens, of qualified equality (Norm 6.2)

If this formula is adopted, it follows that moral justice implies both fairness and qualified equality. Moreover, in view of the Definition 2.5 of a just social structure, or social justice, the above definition entails

COROLLARY 6.4 Moral and social justice are the same.

According to this view the problem of designing a morally just society boils down to a sociotechnical problem, namely that of adjusting the distribution of benefits and burdens over a population in agreement with the available resources as well as with Definition 6.5.

A social distribution $\mathscr{D}$ may be construed as a collection C of items to be distributed, a respect, canon or independent variable V according to which the items are to be distributed, and a function δ, the distribution pattern, relating V to C. In short,

$$\mathscr{D} = \langle C, V, \delta \rangle, \quad \text{with} \quad \delta : V \to C.$$

Moral philosophers and sociotechnologists are interest in two distributions: those of benefits (or rights), and of burdens (or duties). We symbolize them as

$$\mathscr{D}_+ = \langle G, V_+, \delta_+ \rangle \quad \text{and} \quad \mathscr{D}_- = \langle B, V_-, \delta_- \rangle$$

respectively, where G is the total amount of benefits (e.g. goods and fringe benefits), B the total amount of burdens (e.g. tasks, taxes and community services) to be distriubted among the members of a community; V_+ and V_- are the independent variables or canons (such as need, ability, or a combination of the two), and δ_+ and δ_- are the corresponding distribution patterns or curves. To avoid focusing exclusively on either benefits or burdens, we may as well combine the two distributions into the ordered pair $\mathscr{D} = \langle \mathscr{D}_+, \mathscr{D}_- \rangle$. See Figure 6.2. In a just society everyone gets the living minimum required to meet his basic needs, whereas persons with above average ability earn more but may have more burdens than benefits: See Figure 6.2 (*d*).

The ideal of social justice, a powerful political leaven since mid-18th century, has been attacked on pseudopsychological and pseudoeconomic grounds. Thus Freud claimed that the sense of justice, far from being a noble moral feeling, is nothing but an outgrowth of envy and jealousy. And Kaufmann (1973) held that justice is the refuge of the "decidophobe", the heteronomous individual who wished others to make decisions for him — and that the notion of justice is a "decrepit idea" anyway. It goes without saying that neither of the two fantasies has any empirical support, and that neither has been taken seriously by the students of moral psychology: Recall Ch. 5, Sect. 1.1.

As for the pseudoeconomic arguments against the ideal of social justice, they have been put forth by such influential writers as Friedman (1962) and Hayek (1976). They boil down to a seemingly technical,

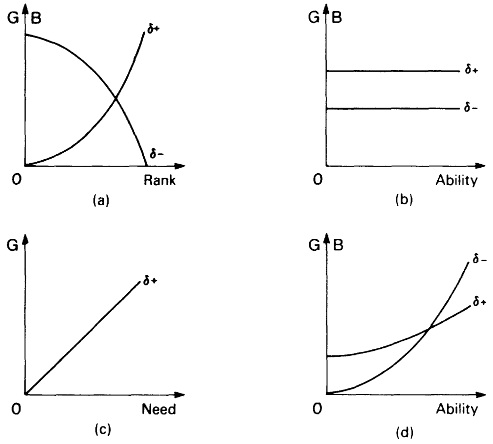


Fig. 6.2. Distributions of goods or rights (G) and burdens or duties (B). (*a*) In an aristocratic society the independent variable V is rank. (*b*) In a radically egalitarian society ability plays no role in distribution. (*c*) Distribution of benefits according to the formula "To each according to his needs". (*d*) Distribution of goods and burdens according to the formulas "Above the living minimum, to each according to his ability and effort", and "From each according to his abilities". $\delta_+ = \delta_-$ for persons with average ability who put in an average effort.

and a seemingly ethico-political objection. The former is that the pursuit of social justice involves governmental regulation and planning, which restrict the free market, which would in turn be the most efficient mechanism for the distribution of goods and services, as competition wipes out the inefficient (rather than the less powerful). There are at least three problems with this claim: (*a*) There are few if any perfectly competitive markets left, and this not only because of governmental intervention in favor of the underprivileged, but also because of the tendency of corporations to merge and expand, because when they go into bankruptcy they ask the state (hence the taxpayer) to bail them out, and because they often ask for subsidies (particularly to agriculture) or tariffs; (*b*) as a matter of fact few markets are in equilibrium, hence efficient: more often than not supply exceeds demand (as in the case of unemployment) or conversely (as in the case of high interest rates); (*c*) the market favors the rich and fails to serve the needier — to the point that most of the poor in Third World countries are marginal to the market and, when they do join it, e.g. to work on cash crops, they only worsen their lot. (See e.g. Moore Lappé & Collins 1988.) In short, the market is neither fully efficient nor just. The problem of devising an economic order both efficient and just will be approached in Ch. 11, Sect. 3.

The ethico-political objection to the ideal of social justice is that the latter is incompatible with individual rights. To a systemist this view is just as mistaken as the orthodox communist opinion that social justice demands the sacrifice of freedom. The two ideas are mistaken because (*a*) the only way of conquering, maintaining or perfecting a just social order is by making use of certain individual rights, such as the right to participate in policy making, and (*b*) the only way individual rights can be protected is in the absence of privilege, i.e. in a society of equals, as we argued in Sect. 2.1. What is true is that (*a*) individual rights are threatened in a society where distributive justice is imposed by force rather than democratically and by self-management, and (*b*) social justice conflicts with private property rights — not with collective property rights though — as well as with the control of political power by the wealthy. (For further arguments in support of the interdependence of social justice, equality and freedom, see Miller 1976, Ackermann 1980, and Nielsen 1984.)

2.5 *Democracy*

The word 'democracy' is usually understood in a strictly political sense, i.e. as the form of government where all the leading political decision makers are elected, and every citizen counts as one elector and has the right to run for elective offices. Being restricted to the political sphere, this kind of democracy is consistent with glaring economic and cultural inequalities, i.e. with social injustice. Moreover it is usually representative rather than participatory, which facilitates the emergence of a class of professional politicians with interests of their own.

However, it is a serious mistake to underrate political democracy just because it is partial. For one thing, political democracy makes it possible to work for social reforms aiming at the extension of democracy to the economic and cultural spheres. For another, empirical research has shown that "Democratic institutions, if maintained for a relatively long time, cause some gradual reduction of income inequality, independent of level of economic development" (Muller 1988). Obviously, the contrapositive too is true: "a high level of inequality reduces a country's years of democratic experience" (ibid.). What holds for economic democratization holds also for cultural democratization: political democracy is a means for both.

The extension of democracy, or self-management, to all the spheres of social life is of course *integral democracy*, which was defined in Vol. 7, Ch. 4, Sect. 4.2 as the freedom to enjoy all of the resources of society, as well as to participate in any social activities, subject only to the limitations imposed by the rights of others. More precisely, we propose to characterize it by means of

DEFINITION 6.6 A society *s* is an *integral democracy* $=_{df}$

(i) *s* is egalitarian (either in the radical or in a qualified manner);

(ii) every member of *s* is treated justly;

(iii) every adult member of *s* participates actively in the management of some economic, cultural or political subsystem of *s*; and

(iv) no subsystem of *s* is managed without popular participation.

In other words, an integral democracy is egalitarian, just, and participatory (or self-managed). Put negatively, in an integral democracy there are (unavoidable and sometimes desirable) inequalities in abilities, benefits, and burdens, but there are no privileges. That is, whatever inequalities there may be are to everyone's benefit, whence, by Definition 6.5, they are not unjust.

We may distinguish four kinds of privilege or unjust inequality, neither of which is supposed to be significant in an integral democracy: biological, economic, cultural, and political. Exceptional physical strength or skill is a privilege only if employed to bully or enslave others, not if put in the service of the weak or of public institutions. Wealth, particularly if inherited, is a privilege because it does not benefit everyone; and it threatens democracy because it enables its possessor to buy political and cultural clout. Cultural prowess is a privilege only if used to exploit or oppress. And the concentration of political power is a privilege because it occurs at the price of the political disenfranchisement of some or even most people; besides, it negates equality and fraternity, and it puts liberty and justice at risk. No such privileges, or injustices, are supposed to occur in an integral democracy. Or, if they do occur, they can be quickly corrected: like science and technology, integral democracy is self-correcting.

True, integral democracy is nowhere in sight. But this does not make it any less desirable. If equality, fraternity, liberty and justice are moral imperatives, as we argued in the preceding sections, so is integral democracy. In fact, equality, fraternity, liberty and justice can only be had and retained in a society where everyone has the right and the duty to participate in the management of the economic, cultural and political resources of his society. In other words, equality, fraternity, liberty, justice and democracy hang together: See Figure 6.1. Hence to praise only one of them amounts to condemning all five.

Integral democracy solves the problem of how best to ensure impartiality in the distribution of benefits and fairness. Rawls (1971) proposed to solve this problem by pretending that people meet in a sort of constituent assembly (the "original position"), where they deliberate and make decisions behind a "veil of ignorance". This veil would guarantee that nobody knows anyone else, whence nobody could foresee what role anyone else is likely to play in society. But this is a misleading fiction because (*a*) society is not a collection of mutually alien individuals but a system held together by common interests and interpersonal bonds of various kinds, and (*b*) no assembly could be convened, let alone be made to work, unless every member of it knew some other members.

Impartiality can only be secured through active popular participation, so that everyone may defend his own interests without infringing on the rights of others. And justice can only be secured by combining popular participation (self-management) with knowledge (not ignorance) of the basic needs, legitimate aspirations, and abilities of the group members. Since such knowledge is only possible in comparatively small groups, integral democracy is necessarily of the grass roots or bottom-up kind. We shall return to this matter in Ch. 11, Sect. 4.

Let us close this section with a historical curiosity. In one of his rare moments of lucidity Hegel (1837 p. 37) proposed a utilitarian justification of participation. Indeed he stated that, if people are to become interested in something, they must be able to participate in it. Regrettably he did not draw the logical consequence, namely the need for participatory democracy.

2.6 *Summary*

Since moral problems can only be had and solved by the living, the most important moral problem of all ages is that of securing survival. Nowadays, to achieve this goal we must begin by realizing that it is morally wrong, as well as stupid, to continue the arms race, to go on polluting the environment, and to maintain the overexploitation of the natural resources, particularly the nonreproducible ones.

Next in line comes the problem of the unequal distribution of benefits and burdens of all kinds: economic, cultural, and political. This is a moral problem as well as a social one if we adhere to the maxim *Enjoy life and help live*, for the plenty of the few derives from the destitution of the many. However, radical or simple equality won't solve the problem, for merit is to be rewarded if we wish to promote individual growth and social progress — which benefits all. The equitable solution is qualified egalitarianism, where the only tolerated inequalities are those that benefit all.

Equality suffices for both fraternity and freedom. On the other hand a divided society is not one of brothers and sisters, and in it only the powerful can make effective use of their rights or freedoms. However, no society can tolerate unlimited freedom. And a good society cannot admit the freedoms to exploit or to oppress, because exploitation and oppression are morally wrong and they lead to justified rebellion.

Equality, fraternity and freedom make justice possible, whereas in their absence justice is hardly possible. Integral justice embraces not only legal justice, or equality before the law: it also involves social or distributive justice. In turn, social justice is nothing but the equitable distribution of benefits and burdens, in such a way that everyone may meet his basic needs and legitimate wants.

Finally, if everyone is to be entitled to pursue her well-being, and is to have the duty of helping others attain or retain that well-being, then democracy is a moral and practical *must*, for in it alone everyone has a say in the management of society. But democracy, to be genuine, must be participatory and integral — economic and cultural as well as political. Purely political democracy can disguise economic or political privilege; purely cultural democracy is compatible with economic exploitation and political oppression; and purely economic democracy is threatened by political authoritarianism and cultural impoverishment. In short, if we wish everyone to be able to enjoy life and help others live, we must redesign and rebuild our societies so that they become integral participatory democracies.

So much for the moral norms we advocate and their ground in the values we support. Our next task will be to compare them to alternative values and morals and to analyze key moral concepts. That is, we now turn from morals to ethics.

PART III

ETHICS

CHAPTER 7

TYPES OF ETHICAL THEORY

The doctrine we have sketched in the previous chapters is of course only one among a large family of ethical theories — or rather more or less articulated doctrines. This family can be partitioned and subdivided in different ways, according to the viewpoint or *principium divisionis* one chooses. Every splitting divides the given family in an exhaustive manner into mutually exclusive classes, such as naturalism, consequentialism, and cognitivism, as well as their contradictories. For example, Hume's moral theory is naturalistic (or humanistic), consequentialist (in particular proto-utilitarian), and non-cognitivist (for being emotivist).

The following partitions are of particular interest to us: The anthropological (model of man), axiological (*summum bonum*), and epistemological (status of moral norms). The first viewpoint leads to the division of moral doctrines into religious and secular. According to the former man is God's creature, the basic moral norms are divine commandments, and the believer abides by them out of love for God — and to secure eternal bliss in the afterlife. On the other hand, according to the secular doctrines man is a natural being left to his own devices and busy pursuing his worldly (but not necessarily egoistic) interests. In turn, the family of secular theories may be divided into subfamilies, such as egoistic and altruistic, teleological and deontological, monistic and pluralistic, and so on.

The plurality of ethical theories, and even of types of such, is a clear indication of the plurality of underlying models of man. The adoption of a scientific model of man, one embracing the biological, psychological, economic, political, and cultural aspects, should help narrow the range of workable ethical theories. On the other hand the adoption of a nonscientific and one-sided model of man — e.g. as being only or primarily *homo oeconomicus*, *politicus*, or *loquens* — is bound to lead to a myopic and impracticable ethical theory, such as Hobbes's or Kant's.

An axiological partition of ethical doctrines focuses on what each of them regards as the supreme good. Table 7.1 exhibits some of them.

From an epistemological viewpoint ethical doctrines can be divided

TABLE 7.1. The primary and secondary values of some ethical doctrines in alphabetical order.

Doctrine	*Primary values*	*Secondary values*
Agathonism	Everybody's well-being	Peace, environmental protection, equality, freedom, justice, democracy
Christianism	God, afterlife	Faith, hope, charity
Contractualism	Individual interests	Liberty, legal justice
Eudemonism	Personal happiness	Justice
Hedonism	Pleasure	Pleasure
Individual utilitarianism	Happiness	Liberty, legal justice
Libertarianism	Liberty	Legal justice
Negative utilitarianism	Individual interests	Not harming
Secular deontologism	Duty to fellow man	Justice
Socialism	Equality	Well-being, community service
Social utilitarianism	Greatest good of greatest number	Liberty, justice
Stoicism	Peace of mind	Devotion to public interest

into cognitivist and noncognitivist. The consequentialist (or teleological) as well as the nonconsequentialist (or deontological) views are cognitivist for they regard moral precepts as pieces of knowledge and therefore learnable and analyzable — and perhaps empirically testable as well. On the other hand moral intuitionism and emotivism, both of which are noncognitivist, have no use for either learning theory or philosophical analysis: The former because it regards moral norms as self-evident, the latter because it takes them to be only expressions of likes and dislikes.

There are two ways of looking at ethical theories: with regard to method and with regard to content. The content of a moral philosophy is of course the collection of moral principles it recommends, whereas its method is the way it proposes to obtain and justify those principles. Though often confused with one another, content and method are mutually separable: It is possible to assent to one of them and not to the other. For example, one may adopt the principle "Thou shalt not kill" without claiming that it was dictated to Moses by Yahweh or trying to defend it alleging that life is sacred. Conversely one may agree with the utilitarians that the worth of moral principles is to be gauged by the consequences brought about by their application, without subscribing to the Principle of Utility.

In the first part of this chapter we shall deal with some questions of method, in the second and third with the content of some influential moral philosophies. We shall have to say something good about nearly every one of them, except for moral nihilism or anarchism ("Anything goes") and naive or radical egoism. However, none of the received views fully satisfies us, this being why we have introduced our own view in the preceding chapters.

1. MATTERS OF METHOD

1.1 *Religious and Secular Doctrines*

Every religion includes a morality but not conversely. In fact, there are non-religious ethical systems, such as Confucianism and Stoicism, which have guided millions of virtuous people over more than two millennia. History shows that religions, though accompanied by morals, have no moral sources: "men have always received their gods from tradition without demanding them to exhibit morality certificates or to guarantee the moral order" (Bergson 1932 p. 217).

The hope of eternal life and the fear of eternal damnation are neither necessary nor sufficient to keep people on the straight path. While they have inspired a few genuine saints, such as Francis of Assisi and Desmond Tutu, they have only too often been used as excuses for killing, torturing or dipossessing infidels, or for defending class privileges or even fueling territorial expansions. The Islamic jihads from Muhammad to Khomeini, the Christian crusades from the 11th century to General Franco's, and the European sacking and oppression of four continents since 1492, are loud and clear testimonials to the potential of religion for evil. Maybe the same atrocities would have been perpetrated without the help of any religion, but the point is that, with but a few individual exceptions, their perpetrators counted on the blessing of organized religions.

There are several reason why religion is not a suitable basis for morality. Firstly there is the problem of logical consistency and empirical support. Take for instance the very concept of God as omnipotent, omniscient and merciful, shared by Christians and Muslims. These three attributes cannot be jointly satisfied. For, if God is indeed omnipotent and omniscient, then he must also be wicked, as only an evil being could have created (or tolerated the emergence of) millions of biospecies only to let them become extinct, as well as remain indifferent to human suffering. If on the other hand God is good and caring, he must have very limited powers as well as very restricted knowledge. The former because he seems to do little if anything to alleviate the misery of his creatures, the latter because he designed them with such serious flaws that no biologist would have tolerated. Even granting that such flaws could have been planted by a possessive god just to try their victims's faith, there is the problem of life after death. For one thing, we have known for quite a while that all mental functions cease on brain death. For another, no dead has yet come back to life to tell us what the afterlife is like.

Secondly, by modern moral standards all religions have serious moral flaws. Thus Hinduism condones the caste system, which is unjust and cruel. So is the dogma of eternal hellfire, because it holds that the damned do not improve by being punished, hence are beyond redemption: They are just as wicked after a million years of hellish torture as before. The Augustinian and Calvinistic doctrine of predestination is even worse, for according to it God himself chooses those who are to be virtuous, rather than electing the virtuous. Even Catholicism and Islam, which promise the reward of virtue, are nothing but variants of egoistic hedonism, for (*a*) holding that virtue, rather than being its own reward, is a means to eternal bliss, and (*b*) not being interested in public morality. (See Albert 1979 pp. 194 ff.)

Thirdly, even assuming that some of the old religions were effective to maintain the cohesion of certain barbarous tribes, they could not possibly have foreseen any of the human survival issues touched on in Ch. 6, Sect. 2.1. In particular human fertility, which was important several thousand years ago, when humans were few and scattered, has become a major threat to the survival of humankind. Therefore in opposing birth control the Catholic Church is worsening a bad situation. This is not to deny that in recent times this same church has worked for world peace and denounced social injustice. (See the encyclical letters "Populorum progressio" of 1967 and "Sollicitudo rei socialis" of 1987.) This departure from tradition is certainly welcome, but it comes in the wake of century-old lay movements and it is neither consistent, radical nor vigorous enough. Nothing short of a Second Reformation could transform Christianity into a good and efficient crusade for solving the current world crisis.

Fourthly, because of its belief in supernatural entities and miracles, and its reliance on revelation and dogma, religion is perpendicular to science, not parallel to it. (Recall Vol. 6, Ch. 14, Sect. 4.2.) Given that scientific research, unlike religious belief, is the intellectual engine of modern culture, as well as the basis of contemporary technology, religion cannot serve as the foundation for anything modern. In particular, it cannot constitute the foundation of any ethical doctrine compatible with science. (For further criticisms of religion from a scientific angle, as well as arguments for a secular ethical system, see Kurtz 1988, the journal *Free Inquiry*, and the publications of the International Humanist and Ethical Union.)

Our society needs a secular morality and a scientific ethics because our moral problems have become so complex that they cannot be solved except with the help of science and technology. Moreover, unlike a primitive tribe, which can make do with a list of don'ts designed to avoid the tribe's breakdown, our society needs a fairly complex ethical system helping us to avoid the extinction of humankind.

In conclusion, it would be dishonest to deny that some religions do contain a few admirable moral precepts, such as the Golden Rule and the injunction to behave kindly toward our neighbors. But any right secular moral code will contain as much and more. And instead of blunting our understanding with bland faith and myth it will stimulate people to argue about moral norms and to seek their improvement in the light of new circumstances and fresh knowledge. Still the religious/secular divide, though crucial from the ontological and epistemological viewpoints, is morally far less important than the egoism/altruism divide. Any religious code leading us to act in an unselfish manner is by far preferable to any secular but selfish moral philosophy — preferable, that is, from a moral point of view.

1.2 *Monism and Pluralism*

It has been proposed that value theories and ethical theories be divided into monistic and pluralistic (Broad 1930, Brandt 1979). Axiological monism holds that there is a single *summum bonum* for everyone, whereas axiological pluralism maintains that there are several independent types of good. Similarly ethical monism lays down a single maximal moral precept, to which all others are subordinated, whereas ethical pluralism holds that there are a number of mutually independent moral norms.

Actually the monism/pluralism distinction does not exhaust the variety of value theories and ethical theories, for there are also hybrid theories. For example, moderate axiological monism — the case of our own axiology — will admit that, although everyone has the same primary and secondary values for having the same basic needs, different people are bound to have different higher order (tertiary, etc.) values, for having different wants or desires. Likewise moderate ethical monism — again, the family to which our own theory belongs — will admit that, although everyone has the same basic rights and duties, hence the same basic moral code, the members of special groups have, in addition, special rights and duties, and therefore special moral norms.

The difference between radical pluralism and moderate monism, whether axiological or ethical, is that whereas pluralism claims that there must be a plurality of mutually independent principles, moderate monism (or moderate pluralism) posits a set of principles united by a single basic principle. See Figure 7.1. In short, we have the following tripartite division of the family of value theories and ethical theories:

Radical monism: A single principle.

Moderate monism (or *moderate pluralism*): Many principles with a single over-arching norm.

Radical pluralism: Many mutually independent principles on the same footing.

For example, utilitarianism is axiologically pluralistic but ethically monistic. Indeed, it posits goods of many kinds, material and cultural, individual and social, but it postulates a single over-arching moral principle, that of utility maximization. This ethical unity encompasses a large axiological diversity, for the utility principle is of the form "For every x, if x is a good, then maximize the utility derivable from x". Likewise our supreme moral maxim, "Enjoy life and help live", sum-

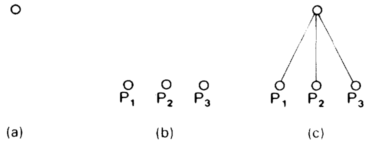


Fig. 7.1. (*a*) Radical monism: A single supreme principle S. (*b*) Radical pluralism: A plurality of mutually independent principles P_1, P_2, P_3, . . . (*c*) Moderate monism (or pluralism): A set of principles subordinated to a single supreme principle S.

marizes an uncounted number of precepts of the form "If you either need or legitimately want x, reach for x; and if others need or legitimately desire y, then help them get y". In both cases the ethical unity covers an axiological diversity.

However, in the case of our ethical theory only the universal part, i.e. the set of norms that everyone ought to abide by, is unitary. Different social groups have different special rights and duties, hence they may legitimately superimpose their own morals on the universal morality as long as they are consistent with the latter. For example, the moral code of the teacher differs somewhat from that of the engineer, not because either of them has special privileges but because each of them has special rights and duties in addition to those common to all citizens.

In each society we have to do then with a whole family of special moral codes, one for each social group, but all of which have a substantial intersection centered in the norm "Enjoy life and help live". See Figure 7.2. Therefore our ethical system is two-tiered: it is *monistic as regards humankind, pluralistic as it regards the various human groups*: youngsters and oldsters, normals and handicapped, manual workers and intellectuals, clerks and managers, and so on. Our philosophy is then moderately monistic (or if preferred moderately pluralistic) both in axiological and ethical matters.

The logical advantage of a radically monistic ethical doctrine is that its consistency is obvious. On the other hand moderate monism and pluralism are bound to have consistency problems in articulating their principles into a coherent whole. However, all ethical doctrines pose consistency problems in application, for none can possibly foresee in all detail the conflicts between the rights and duties of every single person,

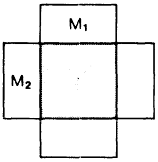


Fig. 7.2. Two codes, M_1 and M_2, with a common core: the central square. Example: M_1 = {⌜Enjoy life and help live⌝, ⌜Be generous as well as demanding with your students⌝}, M_2 = {⌜Enjoy life and help live⌝, ⌜Be helpful and fair as well as strict with your employees⌝}.

or even of every single social group. For example, a teacher's duties may conflict with some of his students's rights or at least wishes; similarly with the physician/patient, manager/employee, civil servant/citizen, and uncounted other pairs. All such conflicts and many more are sources of moral as well as practical problems; and the solution to some of them calls for deliberation and perhaps research too. The fanatic, who worships single-mindedly a single good and a single norm, is spared such moral problems. But he is unlikely to enjoy life and is sure to make other people's lives hard or even impossible. (See Hare 1963.)

Radical axiological and ethical monism are incompatible with individual liberties and a fortiori with democracy, for holding that everyone values or ought to value the same items and behave in the same way. On the other hand radical axiological and ethical pluralism are compatible with individual liberties but not with democracy, for they encourage self-indulgence, which is the enemy of social responsibility, without which democracy cannot work.

In an integral democracy (Ch. 6, Sect. 2.5) every adult person is supposed to be free to choose his values and pursue his own goals as long as he does not stand in the way of other people's basic rights and as long as he performs his basic duties. Since such rights and duties are common to all able-bodied human beings, pluralism must be moderate, not radical, if it is to be compatible with integral democracy.

1.3 *Absolutism and Relativism*

Ethical absolutism holds that moral norms are absolute, i.e. cross-cultural and independent of circumstances. The natural law school, Kant, and the moral intuitionists, such as Moore and Ross, are absolutists. On the other hand ethical relativism holds that moral codes are (*a*) culture-bound and (*b*) functional, i.e. adapted to the needs of the societies that adopt them, whence (*c*) no moral code is superior to any other. Ethical relativism is popular among the functionalist social scientists — whose popularity is declining — and it has been adopted by the philosopher P. Winch (1958) under the guise of "contextual rationality".

There is something to be said for and also against each of these extremes. Firstly, as a matter of anthropological and historical record absolutism is false whereas relativism is true: in fact, value systems and moral codes are culture-bound. Secondly, absolutism is right in holding that certain actions are morally right or wrong in all cultures at all times. Obvious examples of the former: helping children and the weak, and solving conflicts by peaceful means. Obvious examples of universal infamies, even if not always qualified as such: torture, massacring the innocent, and depriving a man of his legitimate means of livelihood.

Absolutism makes any deviation from a given moral tradition into a crime, for regarding that tradition as suprahistorical. On the other hand relativism condones any moral wrongdoing as long as society admits it. Thus if Winch (1958) were right it would be impossible to condemn the Nazis and other mass murderers for they were playing their own "language games" according to rule and in the perspective of their own *Lebensform* or lifestyle.

The root error of ethical absolutism is its belief in an extraordinary source of morals, such as God or infallible intuition, which would render moral norms impregnable to argument and experience, as well as totally independent of extenuating or aggravating circumstances. The root error of ethical relativism is its confusion of morals with *mores*, which confusion may lead to moral turpitude. The absolutist follows blindly the dictates of his moral code without pausing to think of roots or consequences. The relativist follows blindly the dictates of custom without questioning its moral status. The absolutist is a prig, the relativist an opportunist. And both are ethically conservative, though for different reasons.

Our own view is neither absolutist nor relativist but contains elements of both extremes. In fact we share the absolutist's conviction that (*a*) right is right is right even if not so acknowledged in a given society, and wrong is wrong is wrong even if condoned by a given society; (*b*) not all cultures and moral codes are mutually equivalent, whence (*c*) there is objective moral progress in certain periods, and objective moral decadence in others. And we share the relativist's view that morals coevolve with society. But we do not share his belief that every moral code is functional in the sense that it facilitates social life. A moral code obstructs social life if it is obsolete (i.e. if it has not accompanied social evolution) or if it entrenches privilege.

To sum up, we opt for *moderate relativism*: relativism for believing that there are many different and viable moral codes, and moderate for holding that they have a nonempty intersection, i.e. a core of universal or cross-cultural norms.

1.4 *Objectivism and Subjectivism*

Ethical objectivism holds that some moral codes are objective in the same sense as some values are. On the other hand ethical subjectivism maintains that morals are just as subjective as values: that they are either an expression of individual feeling or a set of conventions as groundless as dress conventions. Objectivists believe, and subjectivists disbelieve, that there are moral facts.

An argument for objectivism is that when we argue about certain human actions we judge them to be right or wrong regardless of their agents — though not in independence of every moral code. An argument for subjectivism is that there is no such thing as right or wrong in itself, in total independence of every moral judgment and every moral norm.

We agree with objectivism that there are moral facts, and with subjectivism that these are not in the natural order of things, because morality is an artifact. But so are computers and all other human-made things and processes: they are quite objective without being natural. A moral norm that matches an objective human need or legitimate want is just as objective as the society that adopts or rejects the norm. Such a norm is objective for having biological, psychological or social roots and functions, even though it has no physical existence. A moral norm of this kind is very different from the commandment to worship a deity or to obey a tyrant. Since a moral norm of that type is objective in some sense, so is the (moral) fact it refers to. Thus mutual help is a fact, and so is murder. Only, besides being social facts, the former is right and the latter is wrong: they are facts with a moral status.

The preceding helps answer the question whether moral principles are discovered or invented. Some are invented, though rarely as deliberately as other artifacts. (Recall Ch. 5, Sect. 3.1.) Others are discovered by moral reasoning. Thus once we accept that killing is wrong, we are committed to admitting that legal killing (e.g. capital punishment and military aggression) is wrong as well.

Ethical objectivism is dangerous only if mistaken for absolutism, or the thesis that morals exist in themselves, for absolutism is impregnable to criticism and empirical tests. On the other hand ethical subjectivism is always dangerous, for it entails irrationalism and anarchy (Fishkin 1984). In fact, if moral norms are entirely subjective then it is impossible (*a*) to argue about them, i.e. to propose and evaluate reasons for or against them, and (*b*) to work for a consensus, or at least a substantial majority, to uphold or reform them.

The advantage of (moderate) ethical objectivism over its alternatives can best be appreciated by discussing some examples. Take the Golden

Rule, which comes in at least two versions, one positive, the other negative. The former reads: "Do to men what you wish men to do to you" (*Matthew* 7:12). The negative version reads: "Never do to others what you would not like them to do to you" (*Analects of Confucius* xv. 23). However admirable, neither covers all possible cases, for what one wishes done to oneself may not be agreeable to others, or conversely. A classical example is surgery. I wish to have my hand operated on because it needs repairing, but I do not wish everyone else to undergo this operation, because most people do not need it. Another example is euthanasia. I do not wish to be subjected to euthanasia just yet, but wish that those who clamor for it, for finding life unbearable, be granted their wish. A third example: Drinking pals comply with the rule, for they offer one another drinks that are not good for them. (By the way, a trouble with the Golden Rule in the above versions is that it concerns *wishes* rather than either basic needs or legitimate aspirations. The maxim "Enjoy life and help live" does not have this defect and is far easier to state, understand, and apply. If only for this reason, the Golden Rule cannot — *pace* K. Baier (1966) and others — be taken as the basic principle entailing all of the prescriptions and proscriptions we need. Besides, the rule is highly ambiguous: see Myers 1986.)

In conclusion, we opt for *moderate objectivism*: objectivism because we believe in some objective values, hence in some objective rights and duties, whence in the existence of objectively right and objectively wrong facts; and moderate because of our insistence on *some* rather than *all*. Once again, we have taken a middle road, this time between radical objectivism (the companion of absolutism) and radical subjectivism (the companion of relativism). However, we draw a difference between our rights-and-duties ethics and its rivals, claiming that the former, though neither perfect nor final, is more objective than its rivals for having a firm biological, psychological and social foundation.

1.5 *Emotivism, Intuitionism and Cognitivism*

Ethical cognitivists assert, and noncognitivists deny, that moral principles have a cognitive status ("content") and can therefore be known, analyzed, tested, and altered just like scientific hypotheses. Aristotle, Spinoza, the utilitarians, C. I. Lewis, the present writer and uncounted others have sided with cognitivism. On the other hand Hume, some phenomenologists (e.g. Scheler), neopositivists (e.g. Ayer) and all of the intuitionists (e.g. Ross) belong in the opposite camp.

Noncognitivism is split into emotivism and intuitionism. The former is the view that moral judgments are nothing but expressions of feelings or attitudes, and that their sole function is to goad others to do as we do. (See Stevenson 1946.) We would approve of what we like and disapprove of what we dislike, to the point that one might be tempted to define "right" as that which is generally approved of, and "wrong" as what is generally disapproved of. But this is putting the trailer before the car. In fact, rational people, when given the chance of thinking before acting, start by evaluating a proposal for action, and they do so in light of their value system and moral code as well as in light of the information available to them. They approve or disapprove of the proposal, or of the action, according as it has passed or failed to pass certain moral and prudential tests. Nor is it true that disapproving of something is the same as desiring that it not be done (as proposed by Pollock 1986). Indeed there is such thing as desiring that somebody one hates does what we disapprove of, so that we may be justified in our hatred.

Furthermore, moral norms do not spring directly from moral feelings but are adopted by society, or at least by its leading members, only if they help reinforce behavior of certain types: otherwise they may be proposals but not norms. Since genuine norms rule actual behavior, they are objective even if they are inadequate. And, being objective, they can be investigated empirically. (Recall Sect. 1.5.) As a matter of fact they are being investigated in this manner by psychologists, anthropologists, sociologists, and historians. (Recall Ch. 5.)

This is not to deny that moral judgments have the emotional root and the pragmatic function that emotivists have attributed them: we have admitted as much and more in Ch. 4, Sect. 3. Thus when someone says that he strongly approves of loyalty, he is likely to express a deep feeling and at the same time he attempts to persuade others to be loyal. Without moral feeling we would attend more to the prudential than to the moral. Still, it is one thing to state that a moral judgment or action is motivated by a moral feeling, and another to deny that it can be vindicated or condemned on rational or empirical grounds.

Ethical intuitionism is the view that all moral principles are self-evident and therefore in no need of rational or empirical justification or even examination. Being an epistemological or methodological opinion, ethical intuitionism can be coupled to any substantive moral code, whether egoist or altruist.

Every ethical doctrine that is held dogmatically qualifies as intuitionist. However, there are degrees of dogmatism, hence of dangerousness. The least harmful version of intuitionism is the view that all humans are born with, or come to acquire, the same moral intuitions. This universalist thesis is inconsistent with ordinary experience and with psychology and social science, but at least it is not coupled to an ideology. The most harmful version of intuitionism is the view that in every epoch there is a handful of exceptional individuals, leaders of their herds, who are visited by moral illumination, and whom the masses ought to follow. This elitist intuitionism was actually held by some philonazi philosophers. (See Kolnai 1938.)

We do have moral intuitions, just as we have intuitive knowledge of other kinds. However, they are nothing but our most ingrained moral beliefs: those we have imbibed in our young days. They are so ingrained or internalized that we do not subject them to criticism except at times of crisis. Because moral intuitions are held uncritically, the appeal to moral self-evidence in moral reasoning is irrational and therefore dangerous in practice. In short, moral intuitionism is just as untenable as moral emotivism. (For further criticisms see Bunge 1962b and Brandt 1979.) We must hold on then to moral cognitivism, though without ignoring that there are moral feelings and intuitions.

1.6 *Consequentialism and Deontologism*

A *consequentialist* moralist judges all actions and all rules by their consequences: to him the former are right only if the latter are good. A *deontological* moralist judges a state of affairs good only if it results from a right action. The consequentialist disapproves of rules of conduct that bring about the destruction of his values, whereas the deontologist is ready — or so he claims — to sacrifice his values at the altar of his moral norms. Shorter: Whereas for the consequentialist goodness precedes rightness, for the deontologist the converse relation holds. Hence whereas for the former axiology precedes ethics, for the latter it is the other way round. Catholic and Islamic ethics, as well as utilitarianism and our own system are consequentialist, whereas Stoicism and Kantianism are deontological.

Deontological theories are absolutist and conversely; hence they are open to the same objections that we raised against absolutism in Sect. 1.3. In addition we find the following flaws in it. Firstly, deontologism is inapplicable, for if we are told to further the interests of X we must estimate the possible consequences of our actions on the welfare of X.

It may be argued the Kant himself was a crypto-consequentialist, for by stating the Categorical Imperative (which involves egalitarianism) and sympathizing with the French Revolution, he must have believed that both were good things.

Secondly, if taken literally deontologism is unrealistic and therefore inconsequential for (*a*) overlooking moral feelings and (*b*) ignoring the lesson of learning theory, namely that we learn the basic patterns of social behavior on the strength of the favorable or unfavorable outcomes of our actions rather than by reading Seneca or Kant. (Incidentally, Kant himself was unable to live up to his maxim that we ought to do our moral duty regardless of consequences. For example, he obeyed his king, Friedrich Wilhelm II, who had written him by hand on October 1st, 1794, ordering the Herr Professor to stop distorting and debasing the Christian doctrine if he wished to avoid *unangenehme Verfügungen*. For another, Kant did not recognize his illegitimate child. This is not to be construed as an *argumentum ad hominem* but as evidence for the thesis that deontologism is impracticable.)

Thirdly, deontologism is neither rational nor virtuous: "To do anything except for the sake of *some* expected good would be sheer irrationality or perversity" (Lewis 1969 p. 182). Rational and responsible people do not follow blindly any rules regardless of consequences: they try and foresee and evaluate the latter before taking action, paying attention to their own interests and those of others, to the means available to them, to the opportunity, etc. In short, they combine morality with prudence. Any alternative mode of action is at best inefficient, at worst reckless.

Fourthly, the deontologist is forced to ignore the circumstances of the people he deals with. By rigidly adhering to his maxims he will have not room for compassion or even for extenuating circumstances. He will insist on the cruel dictum *Dura lex, sed lex*. He may not distinguish between deliberate murder and involuntary manslaughter, or between malicious lie and white lie. To him everything is either black or white, whereas the real world is chromatic.

Fifthly, the deontologist is dogmatic, for he has no way of finding out which sorts of action are right and which are wrong except by reference to his own moral code. He must claim that the latter is either a priori (rationalism) or self-evident (intuitionism). By contrast, the consequentialist will ask himself 'Will my principles work in real life? Will they make any difference for better or worse?'. He can use reasoning and experience to correct and update his morals. This is a metaethical argument for consequentialism.

Any of the above reasons should suffice to reject deontologism. However, this is no reason to adopt *naive* or *radical consequentialism*, which holds that the end justifies the means, i.e. that all is well that ends well. Indeed, this short-sighted version of consequentialism is open to at least the following objections. Firstly, naive consequentialism is self-defeating for focusing on the value of the end or goal state, as if the process leading to it did not matter, whereas in fact it does: Recall Ch. 3, Sect. 3.4. (For example, usually terrorism harms national liberation and social justice movements.) Secondly, if an agent employs consistently evil means to attain a good end, he is bound to become callous and corrupt, and even to forget his original goal altogether. (This is the case of the guerrilleros who become bandits.) Thirdly, in certain cases giving up basic principles may be so demoralizing, hence destructive in the long run, that keeping them regardless of consequences may be preferable.

In short, when faced with moral problems we must reckon with principles, real situations and feelings as well as with possible consequences. Shorter: naive or radical consequentialism is inadequate. Since deontologism does not work either, we must opt for a *tertium quid*, namely qualified or moderate consequentialism.

Qualified or *moderate consequentialism* requires that the means employed to attain a goal do not harm others without adequate compensation. Our ethical theory exemplifies this kind of consequentialism, whereas utilitarianism is an instance of naive or radical consequentialism. The difference between the two may be formulated as follows in the case when the (sufficient) means m, the goal g and the side effect s are related by the regularity (law or rule) *If m, then g and s*, and all of these items can be evaluated quantitatively relative to an agent a.

Naive consequentialism (e.g. utilitarianism)

If $m \Rightarrow g \wedge s$ (law or rule) and $V(g, a) > 0$ (value judgment), then $V(m \Rightarrow g \wedge s) > 0$ regardless of the value of m for a or anyone else.

Qualified consequentialism (e.g. our own theory)

If $m \Rightarrow g \wedge s$ (law or rule), $V(g, a) \gg V(m, a), V(s, a)$ (value judgments) and, for every person x other than the agent a, either

(i) $V(m, x), V(g, x), V(s, x) \geqslant 0$, or

(ii) x receives from a a compensation c such that $c \geqslant V(m, x)$, $V(g, x)$, $V(s, x)$, then $V(m \Rightarrow g \wedge s) \geqslant 0$.

The most obvious advantage of consequentialism is that it invites feedback, thus allowing for the updating of information and the correction of principles. Indeed, the consequentialist will arrange his coupling with the system upon which he acts (which may be himself) so as to receive feedback signals informing him about the extent to which his actions have attained the preassigned goal, as well as about the kind and intensity of the (intended or unintended) side effects. See Figure 7.3. And the most obvious advantage of the moderate version of consequentialism is that it avoids Machiavellianism. (For recent discussions of consequentialism see Scheffler Ed. 1988.)

1.7 *Individualism, Holism, Systemism*

Earlier in this Treatise we distinguished three alternative approaches to the study of society: individualism, holism (or collectivism), and systemism (Vol. 4, Ch. 1, Sect. 4.2, and Vol. 7, Ch. 4, Sect. 1.1). Since ethics is about right and wrong social behavior, each of the preceding approaches has its ethical counterpart. The distinguishing features of these approaches can be summarized as follows.

Individualistic or egocentric ethics:

I1 Individuals are or ought to be autonomous, i.e. fully self-determined.

I2 Individual values, particularly liberty, are higher than social values, particularly social justice.

I3 Individual rights take precedence over social duties.

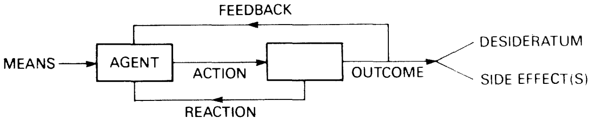


Fig. 7.3. The means-end relation. The agent, if in possession of the suitable means, acts on his target system or patient (possibly himself) causing it to output the desired effect together with some side effect(s). If he is a consequentialist, he monitors the changes in the patient so as to modify his action if necessary. On the other hand the deontologist proceeds headlong without considering the outcome: he passes over the feedback loop.

I4 Society is only a means for the protection of individual rights.

I5 Morals and the law are made for the individual.

I6 Whenever there is a conflict between individual and social values, the former take precedence.

Collectivistic or sociocentric ethics

C1 Individuals are heteronomous, i.e. other- determined.

C2 Social values, particularly stability, are higher than individual values, particularly liberty.

C3 Duties toward social organizations take precedence over duties toward individuals, and these take precedence over individual rights.

C4 The individual is only a means for society.

C5 Morals and the law are made for society.

C6 Whenever there is a conflict between individual and social values, the latter are to prevail.

Systemic ethics

S1 Individuals are in part autonomous, in part heteronomous.

S2 Individual and social values determine each other; in particular, equality and fraternity (mutual help) render liberty and justice possible, and conversely.

S3 Rights imply duties and conversely.

S4 The individual can only exist in society, and the latter should be so designed as to make it possible for everyone to meet her basic needs and legitimate aspirations.

S5 Morals and the law are made for the individual-in-society.

S6 Whenever there is a conflict between individual and social values, the latter are to prevail; however, social organizations should be redesigned so as to diminish the frequency and intensity of such conflicts.

Moral nihilists, hedonists, libertarians and contractualists are ethical individualists. Confucius, Kant, Hegel, Durkheim, and Lenin are ethical collectivists. Aristotle, Spinoza, Hume, Bentham, Rawls and Rescher come close to systemism, for they reject selfishness and admit that the public good takes precedence over individual interests, while emphasizing everyone's right to well-being or even happiness.

Both individualist and collectivist ethics are based on obsolete conceptions of society. The former is scientifically untenable because social scientists do not study individuals but social groups: society is not an aggregate of mutually independent atoms but an "organization of organizations" (Miller 1976 p. 303). Besides, empirical research has refuted the individualist claim that only private interest can motivate involvement in collective action organizations: it has shown that public goods can powerfully motivate such involvement out of feelings of solidarity, altruism, and the sense of duty (Knoke 1988). And the collectivist approach too is inadequate because as a matter of fact social scientists analyze every society into its subsystems, and they take individual needs and interests into account.

Individualists have rightly pointed out that social and ethical holism, or collectivism, crushes the individual. However, they overlook the fact that, wherever individualism prevails, most individuals are crushed by a privileged minority. Moreover, by placing self-interest above the interests of others, individualism is the dissolvent ideology *par excellence*. (A dissolvent ideology is more destructive than a subversive one, for the latter aims at replacing one social order with another, whereas a dissolvent ideology contributes to weakening or even cutting the links or bonds that keep all societies together.)

Individualism is morally corrosive as well as socially dissolvent for, as de Tocqueville (1835) pointed out, it saps the virtues of public life and in the long run it attacks and destroys all others. In fact, by allowing individuals maximal freedom and by minimizing their responsibility to others, individualism tolerates or even encourages the exploitation and oppression of the weaker members of society. Therefore the rise of unbridled individualism in any society is both an indicator and a cause of social disintegration: of the loss of the sense of community and social responsibility. In a cohesive community, where everyone cares for others and gets the help of others, and everyone participates in collective actions, individualism is regarded as sinful.

In short, both individualism and collectivism are scientifically and morally untenable, even though each focuses on one of two real poles: the self and the community. Systemism combines the good and avoids the bad points of those two extremes, for it joins the concern for the individual with that for the social organizations that make individual survival possible and desirable. In the systemist perspective the question 'Which is prior, the individual or society?' is ill-conceived, for there are no such things as the isolated individual "in a pure state of nature" or the society hovering above its members. For the same reason private and social morals, though distinguishable, are inseparable (Ch. 6, Sect. 1.1). Last, but not least, unlike individualism and holism, both of which condone the economic, cultural or political dictatorship of the stronger, systemism goes well with integral (biological, economic, cultural, and political) democracy.

The differences between systemism and its rivals are perhaps best seen when applied to special problems. Take first the autonomy/heteronomy question, with regard to which systemism adopts a middle stand (recall *S1*). Individualism exaggerates autonomy to the point of ignoring that voluntary association is one of the basic constructive human impulses, hence one of the basic human rights. In joining with others an individual not only helps them but also hopes to get benefits resulting from mere togetherness or from collective action, outweighing his partial loss in personal autonomy. In so doing he behaves like a systemist — even while joining the Individualist Party. Moreover, if he joins a democratic organization, the individual has the chance of participating in its decision making process, hence of influencing others while at the same time gaining in social skills. As for the collectivist claim that we are all heteronomous, it ignores the fact that people, especially when young, are normally active, curious, and creative, and sometimes rebellious as well. Only well-orchestrated oppression or utter poverty can render people passive and submissive. In short, normal adults are neither fully self-reliant nor totally dependent.

Our second example is blameworthiness. Individualists hold that only individuals can be blamed or credited for whatever they do, whereas holists blame "the system". Consequently for the individualist moral and social progress or decline depend exclusively on the individual. On the other hand for the collectivist moral and social process or decline are global movements that drag the powerless individual. From a systemist viewpoint individualists are right in blaming or crediting individuals for whatever they are responsible for; but they are wrong in overlooking the fact that every individual is socially conditioned, at least in part, to behave rightly or wrongly. The holist is right in blaming or crediting "the system"; but he cannot possibly reform the latter unless he himself decides to better himself — e.g. to devote more of his time and resources to good social causes. Here as elsewhere the correct answer comes from systemism, which credits or blames the individual-in-society and proposes to reform both the individual and his society through collective action, which is nothing but orchestrated individual action. (We shall deal with collective action in Ch. 10.)

1.8 *Conservatives and Reformists*

Ethical conservatives hang on to traditional values and morals, whereas ethical reformers believe that some values and norms ought to change along with society. For example, ethical conservatives defend the right to procreate at will in an overcrowded world, to own large tracts of land in a country full of landless peasants, and to settle international disputes by the force of arms. Paradoxically, both radical absolutists and radical relativists are conservatives, the former for asserting that values and norms are eternal, the latter for maintaining that every society needs and deserves the morals prevailing in it.

A clear example of an absolutist and conservative ethics is the Christian tract *The Abolition of Man* (1943) by C. S. Lewis, of *Chronicles of Narnia* fame. In defiance of anthropological and historical evidence, Lewis defended a morality, which he called *Tao*, which he regarded as "the sole source of all value judgments. If it is rejected, all value is rejected. [. . .] There never has been, and there never will be, a radically new judgment of value in the history of the world" (p. 29). Actually the *Tao* is a collection of Tory value judgments and moral rules with accent on duty and little regard for rights. It places honor before peace, and charity before mutual help; and, of course, it does not mention the right to enjoy anything or the duty to protect the environment. Notwithstanding its self-proclaimed eternity, this morality was obsolete the moment it was born, or rather rescued from oblivion.

Ethical reformists are motivated by the moral and social ills that plague their society, and by the inadequacy of the prevailing moral and social philosophies to cope with those ills. For example, at the present time they are revolted by their rulers's obsession with military expenditures at the expense of public health and education, and by the corruption that goes with the procurement of defense contracts. Furthermore they believe that something can and ought to be done to rescue humankind from the current social and moral morass.

Ethical reformists hold that morals are made for man, not the other way round, in the sense that rights have to be protected, and duties observed, only as long as they help attain worthy goals by clean means. It follows that ethical doctrines should be flexible, allowing for the changeability of values and norms in the light of new or dwindling resources, new aspirations, and new knowledge. Hence ethical criticism and reform should go hand in hand with social criticism and reform. Whenever either is muffled the other does not take off, and social stagnation, as well as moral corruption or cynicism, prevail.

The practical problem is how to start moral reform. Whereas individualists propose to start by improving ourselves, and holists wait until society changes, systemists suggest that both have to be mobilized at the same time. They believe the historian who tells us that usually moral and social change are initiated by a handful of individuals with new ideas and a strong determination, who succeed in rallying a large enough number of supporters to have the obsolete institutions reformed. In time, the latter will educate more better and happier human beings. We shall come back to this problem in Ch. 11.

1.9 *Summary*

In this section we have examined matters of method, approach or framework, rather than content. In particular we have examined consequentialism, which is compatible with a variety of moral codes, among them all the varieties of utilitarianism. In fact what may be called the *consequentialist principle* states only that everyone should do her best to bring about the best possible outcomes of her action. This principle does not tell us *what* the best consequences or the most adequate means for producing them are. One may choose to fill the schema by specifying that the best consequence is personal happiness or the happiness of the greatest number, eternal life or the glory of God, the honor of mankind or the special interests of a nation, race, class, or party. In short, consequentialism has no special content: it is only a scaffolding for building moral codes that can be evaluated by the consequences of the actions they rule. The same holds for secularism, monism, relativism, objectivism, cognitivism, systemism, reformism, and their rivals: neither is sufficient to formulate moral norms, whether right or wrong ones.

But of course every moral code and every ethical doctrine falls into some of the above categories. Moreover the choice of the correct category should help us evaluate the various ethical *isms* — our task in the next sections. For example, the adoption of consequentialism is necessary (though not sufficient) to avoid dogmatism; of systemism, to navigate between the Scylla of egocentrism and the Charybdis of sociocentrism; of cognitivism, to realize the importance of moral apprenticeship and reform — and so on.

Our own ethical doctrine may be placed in the intersection of the following classes:

(i) *humanism* or *secularism*, for holding that morality is a human creation;

(ii) *moderate monism* (or moderate pluralism) for subordinating all moral principles to a single supreme moral maxim: "Enjoy life and help live";

(iii) *moderate relativism*, for maintaining that whereas some values (e.g. well-being) and some norms (e.g. mutual help) are or ought to be binding and universal, others (e.g. religious faith and worship) are or ought to be optional and group-bound;

(iv) *moderate objectivism* or realism, for asserting that morals are about the real world, and in particular about right and wrong actions, while at the same time only some values, hence only certain rights and duties, are objective;

(v) *cognitivism*, for claiming that the existence of moral feelings and intuitions are not self-justifying;

(vi) *moderate consequentialism*, for holding that actions and norms are to be judged by their consequences, but also by the means;

(vii) *systemism*, for focusing on the individual-in-society rather than on either the individual or society; and

(viii) *reformism*, for holding that much of what goes on in the world is morally wrong, and that we should attempt to reform certain mores in the light of more advanced morals than those upheld by those who have led us to the present quagmire.

2. MORAL *ISMS*: EGOISTIC

2.1 *Nihilism*

Moral nihilism, or amoralism, is the view that the individual need not feel bound to any moral rules: that "everything is possible and nothing matters" (Camus 1951), that "morality is dead" (J. Gilligan 1976), and that "there is no need for morality" (Gauthier 1982). The first quote comes from an existentialist, the second from a psychoanalyst, and the third from a contractualist moral philosopher.

According to existentialism one's responsibility is to oneself: None of us is accountable to anyone else, and none of us has the right to expect anything from others. Psychoanalysis claims to explain morality as a device for sidestepping unconscious guilt feelings stemming from the Oedipus complex. (Girls get their morals, if any, from their fathers.) And contractualism replaces morals with contracts, particularly those entered into in the free market.

Nietzsche (1886, 1888, 1892) is usually credited or blamed for having defended moral nihilism and, in particular, for holding that the search for happiness is contemptible, that the will to power is the highest virtue, and consequently heroism more admirable than goodness. (Not surprisingly Hitler was a great admirer of Nietzsche's: see e.g. Shirer 1960 p. 532.) Actually Nietzsche was preceded by Hegel, who held that might is right, and that "world history is the universal court of law" — i.e. the victorious are always right. However, an even more important source of moral nihilism is the largely unwritten code of conduct of the self-indulging rake, the tycoon, the conqueror, the despot, and the professional criminal: all these practice what the nihilist philosophers preach.

It is instructive to compare Nietzsche's amoralism with the individualistic ethical theories of the Enlightenment which he criticized. Whereas the former despised happiness and the people, the latter believed it possible for individual happiness to contribute to general welfare. Nietzsche extolled the will to power whereas the 18th century moral philosophers sought to ground morality either on the feeling of benevolence or on reason. Whereas Nietzsche's hero is the reckless adventurer or the warrior, that of the Enlightenment is the businessman or the civil servant. Finally, Nietzsche was pessimistic whereas the Enlightenment was optimistic. In short, the former was decadent whereas the latter was progressive.

It is even more instructive to compare Nietzsche's version of moral nihilism with that of Gauthier. Both reject morality for being a constraint on the pursuit of self-interest, but whereas Superman leads a dangerous life Contractman seeks shelter in contract: "The perfectly competitive market offers us a picture of harmony — between individual and society, between freedom and welfare. [. . .] let ourselves be guided by an Invisible Hand [the self-regulating market forces], with no need to submit our backsides to that Visible Foot, morality" (Gauthier 1982 p. 54). We shall analyze contractualism in Sect. 2.4. The point of mentioning it here is to note that, despite the differences between late Romantic hero worship and contemporary market worship, both despise morality and both are fundamentally egocentric.

Moral nihilism is wrong for several reasons. Firstly, normal people have moral feelings, positive such as compassion and negative such as shame, as well as a moral conscience. It takes much mischief to blunt the former and get rid of the latter. Secondly, if we are to live in society we must observe certain norms, among them some basic moral rules, such as abstaining from harming others for fun, and entering into contracts that put life, well-being, or freedom at risk. For this reason moral nihilism, in any of its varieties, is personally and socially destructive. Whoever believes himself to be "beyond good and evil" is beyond the pale.

2.2 *Rational Egoism*

Moral nihilism is only one kind of egoism. Other kinds are individualistic (as opposed to social or universalistic) hedonism, and rational egoism. The former boils down to pleasure-seeking, and the latter to individual utility maximization. The difference between nihilism and its relatives is that, whereas the former rejects all rules, the latter stick to definite rules. And the difference between individualistic hedonism and rational egoism is that the former is purely emotional, whereas the latter purports to involve argument or even calculation.

Individualistic hedonism is nowadays popularly known as "the *Playboy* philosophy", for being promoted by the uninhibited *Playboy* magazine. The gist of this rights-only doctrine is that it identifies the good with the pleasurable. But of course we all know of counterexamples: of good things (like bitter medicines) that are unpleasant, and of bad ones (like *Schadenfreude*) that can cause pleasure. Hence it is not true that everything good is pleasurable and conversely. What is true is that *some* good things are pleasurable and *some* pleasurable things are good. Which are which cannot be told a priori but must be found out by empirical research. In any event, individualistic hedonism can hardly be regarded as a moral philosophy because it ignores duties altogether. Rather, it is straightforward immoralism.

Rational egoism is a far more sophisticated and less destructive, albeit equally wrong, doctrine. It was first explicitly stated by Hobbes (1651) and Spinoza (1670, 1677), though on quite different grounds and with rather different imports. Indeed, Hobbes affirmed, and Spinoza denied, that man is naturally evil, whence that the function of morality as well as of government is to check our naturally base impulses. Whereas according to Hobbes *homo hominis lupus*, for Spinoza *homo hominis res sacra*. The former believed that every one of us is at war with everyone else, whereas Spinoza believed harmony to be possible on the basis of mutual respect. Accordingly, whereas Hobbes's vision of man was somber, that of Spinoza was optimistic. Yet both regarded egoism as viable provided it was combined with reason.

Rational egoism of the Hobbes kind has been vulgarized in our times by Ayn Rand (1964), a sophomoric prophet of the New Right. She held the three cardinal values to be reason, purpose, and self-esteem, and the corresponding cardinal virtues to be rationality, productiveness, and pride. At first blush there is nothing wrong with this simple schema, but trouble starts as soon as its components are analyzed. To begin with, the reason or rationality in question is not that of logic but that of mainstream economy and politology, i.e. the pursuit of self-interest. Such identification of reason with selfishness is one of the most scandalous intellectual swindles of all times. Selfishness is irrational if only because, by killing mutual help and breeding mistrust and hostility, it rips the fabric of any society.

As for purpose or determination, it is worthless except as a means for a good goal. If our goal is evil, we had better be weak in purpose. Self-esteem is destructive unless accompanied by concern for others. (Moreover, it is a psychological truism that an individual despised by everyone, e.g. for being utterly selfish, is unlikely to develop any self-esteem unless he is a psychopath.) Pride is nothing but exaggerated self-esteem, and it is diabolical unless tempered by self-criticism and altruism. Shame, rather than pride, is a powerful motivation of prosocial action, hence of social cohesiveness. (See Darwin 1871.) Finally, productiveness is admirable only if the product is valuable. Being highly productive of offensive weapons of of counter-cultural garbage is a sin not a virtue.

At best rational egoism is irrefutable for making us all selfish. (According to Hobbes even pity is a selfish feeling for being nothing but "fear felt for oneself at the sight of another's distress.") At worst rational egoism is scientifically untenable for the following reasons. Firstly, most people are naturally more cooperative than competitive: they value belonging in a group to the point of being willing to pay the price for it. Secondly, most people are actuated not only by self-interest but also by concern for others, in particular their relatives and friends, but often also for total strangers. Thirdly, some people, particularly businessmen and political activists, take more pleasure in playing the game than in winning: this is why there are at least as many defeats as victories. In short, the thesis that man is basically selfish is not supported by the empirical evidence.

Furthermore, selfishness is impractical and immoral. It is impractical because we need help, hence we must also offer it, for most of our undertakings. And it is immoral because, by definition, there is no morality without a modicum of altruism: The whole point of morality is that it may conflict with prudence. We do not act as moral agents when all we can think of when faced with a moral problem is to ask ourselves 'What's that to me?', 'What's in it for me?', or 'What do I get in return?'.

This is not to say that moral agents must be thoroughly uncalculating. On the contrary, since our resources are limited we must combine moral principles with some calculation if we are to be efficient moral agents. Indeed, every individual has, over a given period of his life, a certain amount R of resources, such as work capacity and time. He can spend R to satisfy his own needs and wants as well as those of his dependents, or in a public-spirited way, i.e. by devoting some of his free time to voluntary work in organizations of public good, or by giving to charities. Call P the total private spending, and S the total social spending over the same period that the available resources R are estimated. Obviously, the sum of P and S equals R, i.e.

$$P + S = R, \tag{7.1}$$

where P includes the individual's savings during the same time. This equation can be represented as a −45° line on the P-S plane. Figure 7.4 exhibits three such lines: for a destitute (D), an average (M), and wealthy (W) individual. (Recall that in this context wealth need not be pecuniary but may be "only" work capacity or time.)

Now, for a self-supporting individual to survive, he must spend a minimum P_0 on himself and his dependents. Hence the most he can devote to public causes is

$$S_{max} = R - P_0. \tag{7.2}$$

Those who devote such maximal resources to the public good are justly regarded as public benefactors or even saints, whereas those who devote nothing to them are regarded as either utterly selfish (if $R \gg P_0$) or as utterly destitute (if $R \leqslant P_0$). Most of us in the

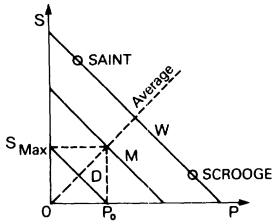


Fig. 7.4. The distribution of an individual's resources R among his personal spending P (including savings in the case of money, and leisure time in the case of time), and his volunteer contribution S to social causes.

industrialized countries fall in between these two extremes. (For more on the reconciliation of morality and prudence see Rescher 1975.)

A common excuse for failing to do one's duty by society — e.g. helping an environmentalist group, or voting — is "It makes no difference whether I do it or not". And a common excuse for taking part in wrong actions, such as breaking the picket line, or joining in the chorus of praise of the despot in charge, is "If I don't do it, someone else will". There are several arguments against using such excuses. Firstly, in many cases we do not know for certain whether or not our act or omission will make a difference. Secondly, personal example is always influential; most people respect those who play fairly without expecting to win. Thirdly, if everyone were to excuse himself from doing his duty by society, there would no society left worth living in. Fourthly, once a person has used an excuse of either type he is likely to continue doing so until he becomes morally degraded. Fifthly, if it takes N persons to prevent an event from causing an amount H of harm, my defection may cause a harm amounting at least to H/N — often more, because my attitude is likely to drag other people as well. (See Glover 1986 for a careful analysis.)

In short, egoism is unrealistic, for most people are not fully egoistic; it is impractical, for the egoist cannot expect help from others; and it is immoral, for without some altruism there is (short-sighted) prudence but no morality proper. (For more see Rescher 1975, Sen 1977.)

2.3 *Libertarianism*

Libertarianism is the view that individual liberty is the *summum* *bonum*. However, the traditional libertarians, such as Mill (1859), were quite different from the contemporary ones, such as Nozick (1974) and Hayek (1976). The former were interested in all the dimensions of liberty: biological, economic, political, and cultural. They were genuine liberals and progressive social reformers in the tradition of the 18th century *philosophes* and the early utilitarians. They favored the equality of races and sexes, economic free enterprise, civil liberties, and cultural freedom. They fought the conservatives of their day. For example, Mill (1873) described himself as a "thorough radical and democrat".

On the other hand most contemporary libertarians only worry about the constraints imposed by the liberal or relief state on property rights. Their chief enemies are the tax collector and the social worker, the planner and the politician. They favor the destruction of the relief state not because it is a mere palliative incapable of effecting social justice, but because it is a burden on the rich. Incidentally, Rawls (1971) is the outstanding exception. Although he too posits liberty as the highest value, he favors the role of the state as a redistributor of wealth. Actually, though variously described as a libertarian, a contractualist, a Kantian, and even a utilitarian, Rawls is neither. Anyway his central interest is not ethics but political and legal philosophy: See Rawls (1985).

We shall center our attention on contemporary libertarianism. Unlike sheer egoism, libertarianism is not a rights-only philosophy. It holds that every right is a constraint that others must obey: it protects the individual from other people's interferences and allows him to do certain things (Nozick 1974). This is correct. The trouble is that, according to the libertarians, the duty to respect other people's rights is our *only* duty. Moreover it would be our duty not to assume any general positive duties, such as helping others, for in observing them we would be bound to violate other people's rights, as is the case with the use of tax revenues to redistribute wealth.

The modern state, be it liberal, socialist, or fascist, presupposes that everyone has certain general positive duties, such as cooperating towards the welfare of others, if not directly at least through paying taxes. (See Garzón Valdés 1986b.) The libertarians reject vehemently the suggestion that we have any positive duties, for any such duty violates some right. (In particular, the duty to pay taxes violates one's right to dispose freely of one's property.) They propose instead the dismantling of the modern Leviathan and its replacement with a "minimal state" composed essentially of law and order forces. In this regard they merely echo what Spencer proposed more than one century ago: they are hardly innovators. In any event they do not propose any *ethical* doctrine, for ethics is concerned with *moral* not just economic or political rights and duties.

However, the worst foe of modern day libertarians is not the hypertrophic relief state, which will be criticized in Ch. 11, Sect. 4, but equality. They hold that equality is incompatible with freedom and, since they put the latter on top, they are inegalitarians. Consequently they are closet authoritarians, for there can be no democracy of unequals anymore than there can be a democratic association of wolves and lambs. When libertarians hold that equality is incompatible with liberty, what they really seem to mean is that *democracy* is incompatible with unrestricted property rights. (See e.g. Dahl 1985.)

Nevertheless there is some truth in the claim that there is some tension between liberty and equality. Firstly, full economic liberty, or rather unbridled free enterprise and free trade, entail the liberty to exploit individuals and entire nations, whose exploitation intensifies individual and national inequalities and restricts individual and national freedom. If the exploited resist, e.g. by joining labor unions or oil cartels, then they curtail the liberty of the exploiters. Secondly, regrettably many egalitarians have been and are politically authoritarian — but so are the libertarians who, in the name of economic liberty, support Third World dictatorships.

The fact of the matter is that (rough) equality is not only compatible with (restricted) liberty but is a necessary condition for the latter: Recall Ch. 6, Sect. 2. If everyone is to have access to the economic and cultural resources of her society, then such resources must not be monopolized by an elite. Put negatively: If economic and cultural liberty are the privilege of a minority, then the greatest number must be deprived of such liberty. Hence genuine liberalism is egalitarian. (See e.g. Ackerman 1980.) Similarly egalitarianism, to be genuine, must not be limited to economic equality, for the moment an egalitarian government monopolizes the economy or the culture it limits equality. In short, we are not faced with a *choice* between liberty and equality, but with the challenge of *joining* them into integral democracy: Recall Ch. 6, Sect. 2.5, and see Ch. 11.

Moreover libertarians have failed to tell us what to do in case of conflict between freedoms, for they extoll liberty in general. What must we do when free enterprise collides with the freedom of access to health care facilities? What when a speaker advocates the derogation of civil liberties? What when the majority rules against minority rights? (See Borovoy 1988.) Such problems cannot be solved if all freedoms are placed on the same level: they call for a ranking based on a definite value system and a definite morality, neither of which has been delivered by libertarianism.

But where libertarians fail most dismally is in addressing the global issues confronted by humankind: nuclear armament, environmental degradation, overpopulation, and poverty. In particular, the rapid deterioration of the environment has left libertarians indifferent (Hospers 1988). And the fact that most people in the world, and one-fifth of the North American population, live below the poverty line, does not bother them either.

In conclusion, libertarianism falls short of a moral philosophy because it does not admit any general positive duties — in particular that of reciprocity — and it ignores all moral feelings — in particular benevolence and compassion. It is only a social philosophy and a very deficient one to boot for being individualistic and for overlooking solidarity, mutual help, and participation, without which no social group is viable. Contemporary libertarians have failed because they have confused freedom with greedom.

2.4 *Contractualism*

Contractualism, or contractarianism, is a distinctive variety of rational egoism. As originally stated by Hobbes (1651) it boiled down to the tenets that (*a*) all humans are basically selfish, hence competitive and even aggressive; (*b*) in the "state of nature" or "original position" human beings were mutually independent and at war with one another, and (*c*) society is a product of a deliberate agreement or contract that reconciles the mutually opposed interests of individuals. Contemporary contractualists, such as Buchanan (1975) and Gauthier (1986), add the following theses: (*d*) an action is right if and only if it conforms to a generally agreed on deal, and (*e*) morals are a byproduct of contracts — i.e. if you keep your contracts, morals will take care of themselves.

The differences between Hobbes and his contemporary heirs are no less interesting that their similarities. Hobbes, and after him Rousseau (1755), held that all men are born equal; on the other hand most present day contractualists are inegalitarian. Whereas Hobbes's chief concern was to protect the individual from himself and attain social and international peace, that of his contemporary followers is to protect the property rights against the liberals and socialists who seek a more uniform or just distribution of wealth. Finally both Hobbes and Rousseau superimposed morals upon the social contract instead of claiming that morality is superfluous. Moreover, Hobbes's ten "Laws of Nature" were at variance with his own view of man as a rugged individualist. Indeed, one of those "laws" (precepts) was "Treat all others as your equals", another "Enjoy in common the things that cannot be divided, and distribute equally the others" (Hobbes 1651 Ch. 15). None of these egalitarian maxims would be acceptable to any contemporary contractualist. (Again, Rawls is the exception, but then he is not a genuine contractualist.)

Contractualism has a grain of truth. It is true that people are *often* actuated by self-interest — though not always; that *sometimes* people engage in rational behavior — but not always; that *some* kinds of social behavior are subject to contract — though by no means all of them; that every contract involves *some* trade-off between rights and duties — though not always an exact balance between them; and that in every society there are *some* principles acceptable to all — though not always sincerely. Contractualism errs in mistaking *some* for *all.*

The psychological, anthropological, sociological and historical theses of contractualism are hopelessly obsolete: Recall Ch. 4, Sect. 3. In particular, Hobbes's and Rousseau's "state of nature" is mythical. So is their hypothesis that the origin of society is a contract. (Forman (1989), a moderate contractualist, concedes both points.) As a matter of fact sociality is an essential property of all humans, and every one of us is born into some society or other. Furthermore primitive man did not contract for social peace: primitive society is typically stable and stateless. The only agreements or contracts that primitive men may have entered into are most likely to have been tacit, and to have involved domestic obligations, division of labor, and trade.

Contractualism does not fit modern societies any more than it does traditional societies. Firstly, the most important human bonds, such as those of friendship, conjugal and parental love, care, solidarity, and social responsibility, are not subject to contract but to feelings and to moral norms. Secondly, both morality and the law make room for a number of categories of sentient being incapable of making any deals, such as minors and the unborn, prison inmates and the very sick, and even domestic animals. Thirdly, neither morals nor the law result from contracts; rather on the contrary, the former are the frameworks without which contracts would be invalid. Fourthly, it is not true that morally right and wrong actions are defined by contract, for many a contract is unfair and therefore can be denounced on moral grounds. Even legal rights are sometimes morally wrong. (Remember that serfdom and slavery, as well as torture and the death penalty, have been fought against on moral grounds.) Fifthly, contracts do not guarantee freedom except in a society of equals. In a stratified society there are power asymmetries that make it possible for people to enter into unfair contracts. (For example, since there is no symmetry between landowner and tenant, or money lender and borrower, the weaker party may have no choice but to accept an unfair deal: he has no bargaining power, hence if he exercises his freedom to decline the deal he and his dependents may starve.) What holds for individuals holds also, *mutatis mutandis*, for nations: The underdeveloped nations are largely at the mercy of the industrialized ones and of the International Monetary Fund, which can stipulate raw materials prices, the kinds of plants to be cultivated, and even political régimes. Genuine and mutually profitable deals can only be struck among equals. Power disparity is the small, usually tacit, print of many a contract. As a wit once said, “The large print giveth, the small print taketh away”.

However, from an ethical viewpoint the main objection to contemporary contractualism is that it misses the whole point of morality. Morality is not about “general agreement” (Scanlon 1982) but about fairness in the allotment of rights and duties. Nor is morality a by-product of agreement, as Gauthier (1986) claims; on the contrary, contracts ought to conform to moral norms. We do not trust our partners for having signed an agreement with us but, on the contrary, we expect them to observe the terms of the contract because we trust their integrity (and their prudence). Only fools sign contracts with crooks. Morality is thus a necessary condition for contracts rather than the other way round. Contractualist ethics stands on its head.

Contracts and legal codes provide neither the fuel nor the moral justification for social behavior. At most they are the rails on which the social machinery is constrained to roll in normal times. (In times of crisis the machine derails or one gets off it before it does.) Moreover, contrary to the individualist ontology underlying contractualism, in civilized society contracts are never person-to-person affairs, for the state is a silent party in any agreement that can be held up in a court of law. For example, no court of law would recognize a contract whereby an employee obliges himself to commit suicide on turning 60, in exchange for a pay rise or for an extra pension for his family. Common decency requires that morals take precedence over contracts rather than the other way round.

Following Hobbes (1651), Gauthier (1986) identifies justice with adherence to the rules of the competitive game, which rules are alleged to prevent mutual destruction. But as matter of fact one of these rules is that, unless protected, the weaker contender loses, perhaps to the point of being left out of the game. In any case, if justice were the same as "the rational disposition to co-operative behavior" in such non-cooperative "games" (Gauthier 1986 p. 115), the very young and the very old, as well as the bed-ridden and the handicapped, would remain outside the realm of justice: being unable to compete or cooperate in a significant manner, they would not be entitled to just treatment. Such behavior would be regarded as grossly immoral even by the primitive.

Real life overflows contracts. By virtue of what agreement does one offer help to someone in distress, or abstain from harming an innocent? What is the contract behind a person's opposition to corporal punishment in schools and prisons? Where is the Invisible Hand that regulates a person's voluntary work in a public good organization? None of such attitudes or actions results *from* agreements, but some of them may result *in* agreements. In fact, moral pressure can change custom or law.

In short, contractualism, particularly in its contemporary version, is scientifically untenable, practically unrealistic, and morally hollow. It has a single virtue, namely that, unlike straightforward egoism, it admits duties. Still, these do not include the moral duties, particularly the supreme duty: Helping to live. If only for this reason contractualism, like libertarianism, does not embrace an ethics.

2.5 *Negative Utilitarianism*

Negative moralities, whether naturalistic like Epicurus's or supernatualistic like that of the Ten Commandments, tell us what is wrong to do, but not what is right to do. Therefore they are guides to inaction not to action.

The simplest negative morality is negative utilitarianism, which enjoins us to minimize suffering -- instead of, say, maximizing happiness. Thinkers as different as Hippocrates and Epicurus, Pareto and

Popper, have espoused the principle *Do no harm.* Hippocrates wished to spare the sick the added misery of painful and ineffective treatments, at a time when hardly any effective ones were know. Epicurus held the principle to prevent the search for one's own happiness from hindering that of others. Pareto wanted to optimize gains in such a way that no party in a business transaction would be made worse off. And Popper's rationale was to prevent the atrocities perpetrated in the name of a future happy world.

Negative utilitarianism restrains egoism but does not encourage altruism. Therefore it is effective when we are in a position to do harm, but not when we are in a position to do good. Furthermore it is open to the following objections. Firstly, it ignores such positive moral feelings as sympathy, empathy, benevolence, and solidarity, any of which may drive us to beneficial action. Secondly, by recommending passivity it condones evil. The spectator who watches impassively a hooligan attacking an old woman, and the citizen who does not bother to vote, comply with negative utilitarianism and thereby tolerate evil. Thirdly, the doctrine discourages people from becoming involved in public affairs. Fourthly, the most expeditious way of implementing the doctrine would be to exterminate humankind, for then human suffering would cease altogether (R. N. Smart 1958).

Negative moralities, utilitarian or not, leave us in the lurch when confronting real life moral problems, for they do not presuppose any value ordering and they fail to suggest what to do. For example, we are told neither to kill nor to lie, which is right in most situations. But what if aborting an embryo will prevent ruining the life of a useful member of society? What if telling a lie will save an innocent life? Negative moralities may tell us that the right path is narrow but they do not indicate where it leads nor what to do in the presence of evil. Hence they are not suitable guides to life. (See more on this in Frondizi 1977.)

In conclusion, negative utilitarianism is a form of egoism. Moreover it is akin to libertarianism, in that both deny that we have general positive duties, hence both condemn any commitment to altruistic causes — which they often ridicule as "bleeding heart do-goodism". Hence they have no claim to moral content.

2.6 *Summary*

We have examined five different versions of egoism: moral nihilism, rational egoism, libertarianism, contractualism, and negative utilitarianism. All five are philosophically unsound for being ontologically individualistic, i.e. for regarding social groups as aggregates of individuals. Consequently they can afford to overlook such systemic or emergent properties as social justice and democracy. And all five enthrone selfishness, though the last three do set some limits on it in order to restrict the damages caused by utterly egoistic, hence antisocial, behavior.

However, limiting egoism is insufficient to get a viable society. Indeed no social group can last long unless prosocial behavior overrides antisocial tendencies. The reason is that nobody is fully self-reliant: Every one of us depends on others, and in a modern society most of these are unknown to us. This mutual dependence calls for mutual help, and mutual help involves caring, sharing, and getting involved in other people's business and, at least occasionally, in public life as well.

Contrary to what Hobbes, Rousseau and his followers think, morality is not just a social control device. It also has a constructive function, that of helping others — some of whom may help us in turn. And moral behavior is partly driven by moral feelings, which are often stronger than self-interest. For these reasons a doctrine that shuns altruism cannot pass for a moral philosophy: the expression 'egoistic ethics' designates a contradiction in terms.

This concludes our examination of the main egoistic doctrines. We now turn to a few genuine ethical doctrines.

3. MORAL *ISMS*: ALTRUISTIC

3.1 *Natural Law*

The natural law schools of moral and legal philosophy hold that all humans are born with certain indisputable rights and duties which would be prior to any convention or legislation. This idea was first explored systematically by the Alexandrian and Roman Stoics. Thomas Aquinas gave it a theological twist by asserting that, though natural, the basic laws are not man-made but a divine gift. The moderns, in particular Grotius, Hobbes, Spinoza, Hume, Montesquieu, Diderot, Rousseau, and Kant, adopted the idea in its secular form.

The most interesting, influential and lasting school of natural law is the original Stoic one. It is naturalistic ("Live according to nature") as well as rationalistic ("Virtue requires knowledge and intelligence"); it combines self-interest ("The ultimate goal is well-being and peace of mind") with public spiritedness ("It is the citizen's duty to participate in politics to promote justice"); it is cosmopolitan (a single world-wide nation) and, in some versions, egalitarian as well. But it is, above all, an upright, nonconsequentialist and austere morality demanding that we live a simple and clean life, that we check our passions, and do our duty. It is a good morality to face scarcity and bear the burdens of life. Since suffering is unavoidable, any reasonable moral code will contain a pinch of Stoicism.

The political fortunes of the natural law idea have been rather strange. Grotius, Hobbes and Kant used it to preach submission to the state. So do some contemporary conservatives, particularly conservative Catholics. But the very same idea of natural rights was used by Montesquieu (1748) and Diderot (1750) to discredit the *ancien régime*, and by the authors of the American Declaration of Independence (1776) and the *Déclaration des droits de l'homme et du citoyen* (1789) to sketch a new society. Thomas Aquinas himself had used it six centuries earlier to argue for some limitations of property rights and to hold that "whatever a man has in superabundance is owed, of natural right, to the poor for their sustenance" (*Summa Theol.* II—II, Question 66, Art. 7).

The German romantics, in particular Hegel and Herder, and later on the socialists Marx and Engels, as well as the historico-cultural (or historicist) school, and much later the legal positivists, took the natural law doctrine to pieces. They pointed out that the concepts of right and duty are not biological but cultural. If a child is born with certain rights it is only by virtue of being born into a society that happens to bestow such rights on him: a different society may not recognize any rights, not even the right to life. This is true: As a matter of historical record, rights are conquered or bestowed, not determined biologically. This holds in particular for property and civil rights, and even more so for animal rights.

Whether moral or legal, norms are artifacts, and as such they can be designed, altered, or discarded to suit the interests of individuals or groups. Hence there is nothing eternal or sacrosanct about them. However, some rights and duties are *basic*, in the sense that without them no individual or social life is possible. (Recall Ch. 4, Sect. 1.) Obvious examples are the right to seek a means of livelihood and the duty to support one's children. These are "natural" only in that they render life possible; yet they have been violated time and again in the course of history. In defending such rights the natural law theorists are right though for the wrong reason.

In short, the natural law school is scientifically indefensible, for the notions of right and duty are moral, sociological and legal, and as such changeable. Still, the universalism and basic egalitarianism inherent in it are valuable and have oftentimes been used to fight economic or political oppression.

3.2 *Kant*

Kant's ethics is an offshoot of the natural law school, hence a mixture of good and bad points. In particular, it is secular and universalistic and, on this count, belongs in the Enlightenment; but at the same time, for placing duty before right, and for defining the good as the observance of duty, it is traditionalist. This ambivalence had a political correlate: Kant was torn between his admiration for the American and French revolutions, and his obligations as a public servant of the despotic Prussian state (Stern 1967 p. 186). Kant inherited this ambiguity from the natural law school, which had been used to defend the *status quo* as well as to criticize it. This and other ambiguities in Kant's work have given rise to a huge Kant scholarship. We do not presume to contribute to it, but shall restrict ourselves to three points of lasting theoretical interest: the right "use" of fellow humans, the alleged primacy of duty, and the requirement of universality.

A basic principle of Kant's is "Treat fellow human beings as ends, not means". This is an admirable norm, particularly in contrast to egoism. However, it needs some amendment for being psychologically and sociologically somewhat unrealistic. In fact, in every social group every human being is at the same time a means as well as an end. Someone who is not a means to anything is totally useless, and no sound person wishes to be useless. In every social group there is a leadership dedicated to attain certain ends with the help of other group members. The point is not to do without leadership but to prevent the selfish or ruthless manipulation of people. In short, the maxim should be "Treat fellow human beings not only as means for the common good but also as ends in themselves, i.e. help them achieve, retain or improve their well-being".

According to Kant duty is absolute and precedes the right, which in turn defines the good. (Consequently ethics precedes axiology.) The practical upshot of this view is the rule "Do your duty regardless of consequences and be saved — or damned". This thesis is open to the following objections. Firstly, it does not match experience, for ordinarily we judge intentions and actions according to their possible consequences; likewise we weigh duties according as they drive to right or wrong actions. For example, a rational person thinks twice before performing a charitable act, knowing that an individual or a community will be better off if helped to learn how to earn a livelihood than if given a hand-out. Kant's demand, of acting out of moral righteousness regardless of motivation or consequence, is unrealistic and it is bound to lead to inefficiency or even to cruelty.

Secondly, Kant's grim conception of morality as a system of inflexible duties to be performed regardless of circumstances and consequences is like the stick without the carrot: it does not make for the happiness of the donkey and it is therefore bound to be inefficient, even for managing a Prussian garrison. An efficient moral code, one embraced willingly, not only constrains but also motivates: it includes not only duties but also rights and ideals, so that it inspires some actions that go beyond the call of duty.

Thirdly, the absence of an axiological foundation makes no room for moral value judgments that are not in the nature of obligations. For example, Kant himself believed that peace is preferable to war, but he had to refrain from stating that it is everyone's duty (including his King's) to favor peace.

Fourthly, Kant's inversion of consequentialism, to read "Good ⇒ Right", condones moral dogmatism and political conformism. Indeed, a Kantian has no way of evaluating an alleged duty other than checking whether it occurs in the socially accepted code. Thus he may approve of the executioner because the latter does his duty. Dogmatism and conformism are morally wrong for being morally insensitive: for not subjecting custom and law to a moral examination, as they must be if we seek to improve ourselves.

All deontological or nonconsequentialist doctrines are inevitably dogmatic, for they require that all moral principles take the form of absolute commandments, of the forms "Do A" or "Don't do A", without qualification or underlying reason. In a consequentialist ethical doctrine moral rules may take the form of hypothetical imperatives of the form "If a is good for b, and you wish b well, and c is likely to bring about a, and it is in your power to do c, then do c". A clear advantage of this approach is that empirical evidence for or against the rule in question can be adduced, and that rational argument about it becomes possible (Harsanyi 1958, Bunge 1960). The possibility of arguing rationally about norms is particularly important in cases of norms conflict. Such conflicts can only be resolved by resorting to the underlying value order (Ch. 4, Sect. 2.2). For example, if the only way to save a life is by lying or stealing, the consequentialist may let the right to life prevail, whereas the Kantian won't: he will stick to the inhumane *Fiat justitia, pereat mundus.*

Having indulged in some Kant-bashing let us close with a note in praise of the famous Categorical Imperative, i.e. the meta-rule that enjoins us to act only according to rules that are universal or universalizable, i.e. valid for all humans. (For recent discussions of this matter see Potter & Timmons Eds. 1985.) This norm presupposes the principle of the equivalence of all human beings, common to the Stoics and the early Christians, the utilitarians and the socialists. It is still a task before us to elaborate a minimal moral code that may be embraced by everyone. (We shall return to this in Ch. 12, Sect. 3.) Incidentally, note that the Categorical Imperative is not a norm, because it does not concern behavior in a direct manner: it is a metanorm, or ethical postulate, for referring to rules. See Figure 7.5.

In short, Kant's ethics contains some admirable principles, but it is flawed on three counts: it is short on rights and long on duties; it places rules above people and their circumstances; and it contains too few substantive principles. Yet because of its worldliness, altruism, universalism, and egalitarianism, it continues to be a source of ethical inspiration in addition to being an academic industry.

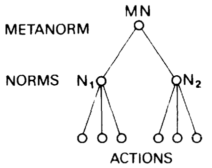


Fig. 7.5. Norms N_1 and N_2, each of them referring to a set of actions, are in turn ruled by metanorm MN.

3.3 *Utilitarianism*

Utilitarianism is the collection of ethical doctrines that measure the worth of an action by the utility it procures, i.e. by the value of its outcome. Utilitarianism, a typically consequentialist and monistic doctrine, goes back to Hume (1751).

From a substantive viewpoint we may distinguish two kinds of utilitarians: the individualistic or egoistic utilitarians and the universalistic or altruistic ones (Sidgwick 1907). The former are followers of rational egoism (Sect. 2.2). Egoistic utilitarianism is the calculating selfishness inherent in mainstream economics and politology. Its slogan is "Maximize your expected utility". It consecrates greed, under the name of enlightened self-interest, as the greatest virtue.

Classical utilitarianism is of the universalistic or altruistic kind. It holds, following the materialist philosophers Helvétius (1758) and Priestley (1768), that one should always pursue "the greatest happiness of the greatest number". This deceivingly simple and attractive formula was adopted by Bentham (1789). So as not to leave any doubts concerning the social character of his doctrine, he conjoined it with the egalitarian principle "Each is to count for one and no one for more than one". In short, classical utilitarianism, from Bentham and Mill to Harsanyi (1985) and Smart (1973), proposes that we maximize the aggregate or social utility or welfare — whatever this may be — rather than the individual one.

However, the classical utilitarians took it for granted that every benevolent and reasonably smart individual would automatically maximize social benefits by pursuing his own personal interests. Mill (1863) went so far as to attempt to formally deduce universalistic from hedonistic morality. He did not realize that this could only be done by explicitly adding some such principle as Bentham's egalitarian maxim, or Kropotkin's mutual help precept. For, if every individual is allowed maximum liberty, then destructive competition, exploitation and even oppression are likely to ensue. (Recall Sect. 2.3.)

The combination of the utility principle with egalitarianism solves moral problems of the kind formulated by Harris (1975). Two patients are bound to die unless one of them gets a new heart and the other a new lung. There are no such organs in stock. The patients propose that a healthy man be killed, so that the two organs will become available. The egoistic utilitarian would agree to such a crime because it would double the total amount of happiness. On the other hand the altruistic utilitarian will count no man as more than one, hence he will prevent the crime. But, being an activist rather than a quietist, he will publicize the need for the organs in question, and may thus save the two patients.

Bentham and Mill were social reformers. On the other hand Sidgwick (1907), though a utilitarian, criticized the attempt to break the moral code accepted by his society: he was a conservative. (This suggests, incidentally, that utilitarianism has no moral code of its own.) Others have been revolutionaries. Lenin (1920 p. 669) is a case in point, for he held that "morality is what serves to destroy the old exploiting society and to unite all the toilers around the proletariat, which is creating a new, Communist society". This statement is utilitarian because (*a*) it presupposes Marx's thesis that the transition from capitalism to communism would benefit everyone, and (*b*) it draws no difference between morally right and wrong means, but adopts Machiavelli's maxim "The end justifies the means". (For more on Marxist ethics see Kautsky 1906 and Hörz and Wilke Eds. 1986. Incidentally, Marx had criticized capitalism chiefly on moral grounds: for being exploitative and for putting a price on everything, hence for being morally corrupting. And, while Lenin criticized "bourgeois morality", he was quite puritanical; in particular, he campaigned against free love.) In short, utilitarianism is morally and politically uncommitted.

So much for substantive issues. Let us now deal with a methodological problem. Brandt (1959) distinguishes two main varieties of utilitarianism: *act utilitarianism* and *rule utilitarianism*. According to the former an act is morally right only if it has maximally beneficial consequences: it is an opportunistic doctrine rather than a principled one. On the other hand, according to rule utilitarianism (*a*) an act is morally right only if it abides by some moral rule, and in turn (*b*) a rule is morally right only if the acts it covers are maximally beneficial. The jurist J. Austin (1832) put it in a nutshell: "Our rules would be fashioned on utility; our conduct, on our rules". Harsanyi (1985 p. 44) puts it more explicitly: "the very purpose of rule utilitarianism is to identify the moral code that would maximize expected social utility if it became the accepted moral code of society".

In principle, rule utilitarianism is the more inclusive of the two varieties of utilitarianism, for it takes the utilitarian maxim and builds upon it. Whereas act utilitarianism is only a criterion for decision making, rule utilitarianism is both a criterion and a code (Ratner 1984 p. 755). Hence *prima facie* rule utilitarianism is vastly superior to act utilitarianism. But the difference is illusory because rule utilitarians have not proposed any rules of their own: Like so many other contemporary moral philosophers, they have been so busy doing metaethics that they have had little time to spare on morality. The upshot is that rule utilitarianism can only be put to work if conjoined with a moral code external to utilitarianism. Usually this code is the received one.

Harsanyi (1985) has unwittingly supplied a clear example exhibiting the moral emptiness of rule utilitarianism. Marital unfaithfulness, if kept secret, maximizes the aggregate utility of all three (or more) people involved. But if discovered it would not only cause misery to all the individuals concerned, but would also weaken trust in matrimony and, in general, in the prevailing moral code. To avoid the devastating consequences of act utilitarianism the rule utilitarian invokes the standard morality, which happens not to be utilitarian. In short, rule utilitarianism has so far only made a methodological contribution by showing the weakness of act utilitarianism. (For another telling example showing the need for extra precepts see Gibbard 1984 p. 280.)

Smart (1973), an act utilitarian, claims that rule utilitarianism encourages a rigid and superstitious attitude towards moral rules — what he calls 'rule worship'. But the total absence of rules is confusing and paralyzing. We should object to bad rules not to rules in themselves. True, abiding dogmatically by a rule can have damaging consequences; but this holds in every field, not only in ethics. The only effective way of fighting dogmatism, whether in ethics or elsewhere, is by reviewing the principles once in a while: by examining them conceptually and in the light of experience, i.e. by their practical fruits. But we must have rules before we test them, and so far rule utilitarianism is only the name of a project.

Let us now evaluate utilitarianism of the universalistic or altruistic kind. First the credit side. It has, from our viewpoint, a number of virtues: (*a*) it is secular rather than theological, hence it harmonizes with modern culture better than traditional deontologism; (*b*) it is consequentialist, hence less dogmatic than deontologism; (*c*) it is activist rather than contemplative; (*d*) it involves a shift from medieval values, such as piety, self-denial, entitlement, position, honor, and dignity, to down-to-earth usefulness, and (*e*) it is socially oriented — whence it is unfair to blame it for the wars, massacres and persecutions of our time, as Hampshire (1983 pp. 84—85) has done.

Now for the debit side. An obvious flaw of utilitarianism is that it is not psychologically true: We act mostly on impulse, from empathy or from habit, rather than on calculation. For this reason Darwin (1871 p. 162) stated that the utilitarian maxim is to be regarded as the *standard* but not as the *motive* of conduct. (For more on the neglect of moral sentiments on the part of most utilitarians, see Rescher 1975.) But even if utilitarianism is regarded as a normative not a descriptive view, it is open to the following objections.

Firstly, utilitarianism has no underlying value system aside from the crude dichotomy of all objects into useful and harmful. It tells us neither *what* is good, hence desirable, nor *what* is morally right, hence doable. It does not even distinguish needs from wants. Hence it does not help us evaluate plans of action. Not surprisingly, we find self-styled utilitarians on both sides of nearly every issue, e.g. among the advocates and the critics of nuclear deterrence and first strike. (See e.g. Hardin *et al.* Eds 1984.)

Secondly, utilitarianism, though centered on the utility concept, lacks a correct and applicable utility theory. In fact, the existing utility theories and their applications, particularly game theory, involve utilities (subjective values) that are neither interpersonally comparable nor subject to general laws nor — and this is fatal to any talk of aggregate utility — additive. (Recall Ch. 2, Sect. 1.1 and Ch. 3, Sect. 3.3.) For this reason some utilitarians resort to preferences — which, being often irrational, do not constitute a safe basis for either efficiency or justice. Others resort to phony numerical assignments, particularly when using game theory. For example, in defending cooperative utilitarianism, Regan (1980 p. 124) sets up a social payoff matrix the entries of which, pulled out of a hat, presuppose what is to be proved, namely that cooperation pays more than defection.

Thirdly, utilitarianism, particularly of the act type, focuses on goals and overlooks the problem of means. Consequently (*a*) it may condone the maxim that the end justifies the means, and (*b*) it does not help us to choose the (prudentially or morally) right means. However, utilitarianism may occasionally help us *reject* certain means. For example, a utilitarian criminologist will oppose torture and the death penalty by arguing that (*a*) they are inefficient as crime deterrents and (*b*) they hurt not only their targets but also their executors for troubling their moral conscience or blunting their moral sensibilities.

Fourthly, for insisting on the maximization of aggregate or total utility, utilitarianism skirts the problem of distributive justice (Rescher

1968, Sen 1970, Rawls 1971). An example taken from Sen (1970 p. 143) will clarify this point. Consider a society composed of two units (individuals or social groups) *A* and *B*, and two possible social states, *S* and *T*. In state *S*, the welfare of *A* is 1 and that of *B* is 0; in state *T*, the welfare of *A* and *B* are the same, namely 1/2. Thus the expected (average) utilities of the two social states are the same, i.e. 1/2; but whereas state *S* is grossly unfair to *B*, state *T* is fair to both parties. Still, this defect can easily be repaired by resorting to Bentham's egalitarian principle, "Each is to count for one and no one for more than one". The resulting distributive principle would be: "Of two social states, prefer that which both maximizes total welfare and distributes the latter in the fairest fashion". Regrettably this principle is vague because of the inexactness of the expression 'total welfare'.

Fifthly, utilitarians ignore the complex process of competition and cooperation among individual or collective agents and, in particular, the bargaining processes and the resulting contracts (Gauthier 1986 p. 127).

Sixthly, the maximization requirement faces the following difficulties. To begin with, it can rarely be put into practice because it is exceptional that two mutually independent functions may attain their maxima at the same time (Mosteller 1981). Further, maxima are not necessarily optima. We wish to maximize our chances of *attaining* our goals, but the goals themselves need not be maxima. Think of the harvesting of a natural resource: A rational manager will exploit it at such a rate that it will be profitable without destroying the resource, whereas the maximizer is bound to kill the goose that laid the golden egg — as is currently the case with fishing, logging, and even cultivation.

In addition to these genuine difficulties a number of philosophers have found imaginary flaws in utilitarianism. For example, it has been said that it constrains the realization of an individual's personal (or life) project. This is true, but it applies to every ethical doctrine except straight egoism, which is immoral: no personal project should involve jeopardizing other people's well-being. It has also been claimed that the classical utilitarian principle allows conflicts between the greater good and the greater number: that whereas some might wish to maximize the total amount of goods, others prefer an equal distribution of a lesser bounty. But this is an imaginary conflict, for it is technically possible to increase production and enhance fairness at the same time. What is true is that certain extremist ideologists would have us believe that we must opt for either goal. However, none of them is a utilitarian in the tradition of Bentham and Mill.

In short, utilitarianism is deficient rather than basically wrong. It lacks an underlying value system, it is not supported by a true value theory, and it has no clear notion of aggregate or social utility. Worse, utilitarianism draws no distinction between the moral and the prudential. This is why it contains no moral code, in particular no norms concerning the rights and duties that arise from basic needs or legitimate aspirations. Like Kantianism and contractualism, it is morally empty. Regrettably utilitarianism is often rejected for the wrong reasons, mainly for being consequentialist or for being concerned with social welfare.

3.4 *Agathonism*

Our own ethical theory, formulated in Chapters 4 to 6 on the basis of the value theory defended in Chapter 1 to 3, may be called *agathonism* (from άγαθόν = good), for favoring the pursuit of the good in both the personal and the social spheres. There is no point in summarizing it here, but in may be worthwhile to note some of its similarities and differences with alternative doctrines.

Agathonism shares several important features with rival ethical theories. In particular it is altruistic, basically egalitarian, universalistic, and consequentialist. Furthermore it agrees with the natural law school in that some rights and duties are natural because they derive from biological needs. It shares with deontologism a concern for duties and therefore the rejection of rights-only doctrines. And it agrees with utilitarianism in assigning no value, moral or otherwise, to any actions unlikely to enhance personal well-being or social welfare.

On the other hand agathonism differs from most of its rivals in that it rests on a definite value system and a definite value theory. Moreover, far from being vague platonic ideas, its values are primarily rooted to basic biological, psychological or social needs, and secondarily rooted to legitimate aspirations. The top value is the survival of humankind, which cannot be secured except in peace and in a favorable environment: See Figure 12.1. (These two conditions are jointly necessary but mutually independent.) The moral correlate of the value judgment *Primum vivere* is the precept *Enjoy life and help live*, which combines self-interest with social responsibility. Every legitimate human right and every genuine human duty is subsidiary to the maximal norm.

Agathonism has the philosophical advantages noted in Sect. 1.9. It has an additional advantage that will become apparent in Part IV of this book, namely that it underlies a realist theory of action and a theory of integral democracy and integral national and international development. Moreover we submit that agathonism satisfies all of the desiderata listed in Sect. 3 of the Introduction. The reader will judge whether this claim is justified.

3.5 *Summary*

We have reviewed four altruistic ethical theories: natural law, Kantianism — a variant of the former —, utilitarianism, and agathonism. We have found merits in the first three, mainly their being altruistic, universalistic, and basically egalitarian. But we have also found them wanting on several counts.

For one thing, none of the first three theories is realistic enough: hardly any of them can be practised consistently by real people. For example, it is simply not true that all rights and duties are thoroughly natural and permanent: most are society-bound and therefore historically changeable. Kant's ethics lacks an underlying value theory, it inverts the relation between the good and the right, and it dwarfs rights by comparison with duties. Utilitarianism too is unrealistic for overlooking moral sentiments and overrating the extent to which we calculate expected utilities. As well, it lacks an adequate value theory and, what it worse, it does not contain a moral code, but only the promise of one.

Agathonism lacks the shortcomings of its rivals and it has several virtues which none of its rivals has. One of them is the philosophical advantage of harmonizing with a realistic systemic social philosophy, which avoids the myths of both individualism and holism. Another is the practical advantage that it addresses genuine and pressing problems, first of all that of the survival of the human species. Where other ethical theories emphasize pleasure, utility, freedom, justice, or equality, ours puts the accent on survival and its two chief conditions: peace and environmental protection.

CHAPTER 8

ETHICS *ET ALIA*

Ethics is a branch of knowledge that has a direct bearing on action. Hence it is closely related to other branches of knowledge and to all fields of human action. All this may sound platitudinous, but it has been denied by various schools of thought. For example, the emotivist moral philosophers (Ch. 7, Sect. 1.5) claim that ethics has no cognitive content. And many a businessman, as well as many a politician, denies that his actions ought to be constrained by moral precepts. Moreover, there is an entire philosophical school, namely contemporary contractualism, which denounces ethics as a straitjacket or at best as a byproduct of business transactions. (Recall Ch. 7, Sect. 2.4.) Hence the examination of the relations between ethics and its theoretical and practical relatives may not be redundant.

Ethics used to be defined as "practical philosophy". However, until recently modern moral philosophers seldom dealt with practical problems. They limited themselves to stating empty formulas such as "Do what is right", without telling us what is the right thing to do. Or, in the wake of the so-called linguistic turn of philosophy, they confined themselves to analyzing some ethical concepts, norms, and doctrines. In either case they avoided coming to grips with burning and pressing practical problems demanding the taking of sides, e.g. for or against social justice, for or against nuclear disarmament, and for or against the use of violence.

This purely academic orientation of moral philosophy has changed radically over the past two decades. Today applied ethics, rather than theoretical ethics, is the rage, as can be seen by perusing such journals as *Ethics, Journal of Applied Philosophy, Philosophy and Public Affairs, Public Affairs Quarterly, Environmental Ethics, Philosophy and Medicine*, or the *Journal of Business Ethics*, not to mention an amazingly large output of scholarly books on real life moral problems, particularly in the areas of bioethics, political philosophy and legal philosophy. All this is to the good, for an ethics divorced from moral problems is, just like an ontology and an epistemology alien to science and technology, mere intellectual gymnastics. However, concern for specific practical problems should not make us forget theoretical problems, for applying a discipline to investigating a problem presupposes that the discipline is available — and, as everyone is likely to agree, theoretical ethics is not in good health. Like any other branch of philosophy, ethics does not thrive except when in close contact with other research fields as well as with genuine problems.

In this chapter we shall look into the relations between ethics and other branches of knowledge, from logic to technology. We shall also examine the relations between ethics and certain kinds of action, from defending rights to saving lives, and from doing business to waging war.

1. ETHICS AND KNOWLEDGE

1.1 *Ethics and Logic*

The emotivists subscribe to Hume's dictum: "Reason is, and ought only to be, the slave of the passions". More explicitly, reason should only serve to devise the best means to ends that are themselves beyond reason, i.e. that are opaque to analysis and impregnable to criticism. It must be admitted that Hume was partly right: Certain "passions" are not to be reasoned out. These are the basic needs or drives, those which must be met if we are to survive. And there is the passion to know, which in the case of basic science is not the slave of any other passion.

On the other hand there are "passions" that must be controlled for the sake of either prudence or morality: They are our wants, desires, or wishes. In particular, we must examine and check those wants the pursuit of which may endanger lives or put the well-being of others at risk. As a matter of fact we do so all the time. For example, most of us do not live to eat or drink, are not promiscuous, do not grab everything we like, and do not indulge whatever dream of domination we may have. This control is partly prudential and partly moral, and in either case it involves some arguments or reasonings. Hence we must alter Hume's dictum to read: "Reason is, and ought to be, *inter alia*, the slave of our basic needs and legitimate aspirations, as well as the master of our wants."

We submit that the role of reason is the same in morals and in the theory of morals as in science or in technology: In all four cases reason is necessary to formulate problems, to state, articulate and evaluate principles, to construct and evaluate arguments, and to examine conceptual systems (e.g. codes of conduct) as to their internal and mutual consistency. But, just as in science, in morals and in ethics reason is insufficient to discover or invent as well as to evaluate principles. In both cases, unless we are dogmatists, we shall resort to experience as well. Indeed, reasonable, experienced and morally sensitive people do not face moral problems as either pure empiricists or pure rationalists. Rather, they behave as ratioempiricists. (For this notion see Vol. 6, Ch. 15, Sect. 2.1.)

To be sure, in order to have basic experiences of good or right we need not engage in reasoning, in particular in calculation. But we do need straight thinking when deliberating about complex choices. "Since thinking and believing must be involved in any decision of doing, the logical validities are a *necessary* condition of any principles of validity of prudence or of morals" (Lewis 1969 p. 69). The irrationalist rejection of reasoning in the realm of morals is bound to lead us to dogmatism, inefficiency, and even evil. This is not to say that reason suffices: it is hard if not impossible to practice morality without moral sentiments and without willingness to compromise on secondary matters. In practical matters it is often better to be reasonable, wise and compassionate, than coldly rational, very well informed, and intransigent.

The role of reason in morals is not confined to the choice of means: It should also control that of goals, particularly since in life goals often get transmuted into means and conversely. For example, sometimes we want freedom in order to learn, whereas at other times we want to learn in order to free ourselves from some yoke. Yet, some philosophers have held that we should shun reason in the moral sphere because it makes no room for moral choice. This view is mistaken: Rationality clashes with arbitrariness and with groundless convention, not with moral autonomy. An autonomous — or, to be more realistic, semiautonomous — agent can act either rationally or irrationally, depending on her background and on the situation. Rationality (R) and autonomy (A) are mutually independent variables. Hence all four combinations are possible: AR (free agent observes rational rules), $A\bar{R}$ (free agent acts irrationally), $\bar{A}R$ (Slave behaves rationally), and $\bar{A}\bar{R}$ (slave violates rational rules).

Any moral problem, i.e. any conflict between values or the corresponding norms, calls for some ratiocination. For example, if two values or norms come into conflict, we place them in our value system or in our moral code, and pick the highest of the two. In this case pigeon-holing and comparison solve the problem. The solution of other moral problems involves much more complex reasonings – particularly when we try to extricate ourselves from a moral predicament.

Given any moral or ethical argument, in principle it is possible to construct a counterargument. We distinguish three kinds of dispute over moral arguments: *Type I* controversies, which bear on the facts of the matter (i.e. the situation of the people involved), *Type II*, which concern premises (moral or prudential principles), and *Type III*, which are about inferences.

Type I disputes (over the evidence) are in principle the easiest to settle, unless of course one of the parties has no access to the facts, or refuses to find them out. (Until recently this was the case with the disputes over the manufacture of nuclear weapons.)

Type II disputes (over premises) are the toughest because the choice of principles may be uncritical or it may be tainted by self-interest. For example, the alleged right to first nuclear strike is based on the principle that national security comes before anything else — and if you do not believe this then you are a traitor. Those who dispute the alleged right argue from a different principle, namely that the survival of humankind comes before national security. (They may also add the prudential principle "The more weapons, the higher the risk of war, hence the smaller the security.") When principles clash, only an appeal to their possible consequences may alter deep-seated convictions or even the perception of self-interest.

Type III disputes (over inferences) would seem to be the easiest to settle, but this might be a superficial impression. Do not moral arguments involve commandments, which are neither true nor false? If they do, then moral arguments do not satisfy ordinary logic, which deals with (true or false) statements, but call for a special logic, capable of validating or invalidating inferences involving imperatives, whether categorical such as 'Do not do *a*' or hypothetical such as 'If *a* is the case, do *b*'. As a matter of fact many philosophers have attempted to build a logic of commands or imperatives, i.e. a *deontic* logic or logic of *ought* as opposed to the ordinary logic of *is*. But to this day, after decades of ingenious efforts, no satisfactory system of deontic logic has emerged (Weinberger 1985). In particular, any deontic logic with an underlying modal logic will include Aristotle's law "If *a* is the case, then *a* is possible". This principle, correct in the case of real or physical possibility, becomes a moral or legal monstrosity if "possible" is interpreted as "permissible". (For the utter uselessness of modal logic in philosophy, see Vol. 3, Ch. 4, Sect. 1.2.) We shall return to these matters in Ch. 9, Sect. 1.4.

Actually there is no need at all for a special logic of commands or imperatives, and this is for the following reason. Although commands are grammatically and pragmatically different from norms cast in the propositional form, they are conceptually (or semantically) equivalent to them. For example, "Do not teach lies" is conceptually (though neither linguistically nor psychologically) equivalent to "It is wrong to teach lies", or "It is not morally permissible to teach lies". In our system (Chapters 4 to 6) all principles are propositions or easily transformed into such. For example, our supreme principle, "Enjoy life and help live", abbreviates the proposition "Everyone has the right to enjoy life and the duty to help others live". (More on the propositionalization of commands in Ch. 9, Sect. 1.5.)

What holds for moral norms holds for inferences involving such norms. For example, the argument

All revenge is wrong.	$\forall x\,(Rx \Rightarrow Wx)$
Action *a* is revengeful.	Ra
∴ Action *a* is wrong.	$\therefore\ Wa$

is an ordinary deductive argument. In conclusion, Type III disputes over moral matters can be settled by resorting to ordinary logic. The moral is: Do not be misled by linguistic wrappings.

Like in other fields, in ethics we may wish to *explain* actions and norms. For example, if we are asked why it was wrong to assassinate *X*, we may respond: 'Because all murder in cold blood is morally wrong'. And if in turn we are asked to explain this norm, we may give a number of replies, e.g.: 'Because everyone has a right to life', 'Because murder debilitates the social bonds', 'Because one murder may incite another', etc.

In addition to explaining actions and norms, we may also wish to *justify* them, i.e. to show that they are morally (or legally) right or just. Note the differences between justification and explanation: (*a*) only human actions are justifiable or unjustifiable; (*b*) we may explain unjustifiable actions, and justify unexplained ones. We justify actions in terms of norms and circumstances. (Excuses are incorrect justifications when they involve either lies or the opportunistic discarding of some moral principles.) Unlike deontologists, who feel no need to justify their principles, consequentialists do feel such a need. They are committed to showing that their principles are necessary to realize their values. But this is a matter for the next section.

Finally, note that not everyone is eager to engage in moral arguments, particularly if they bear on moral principles. There are several tactics for stopping moral argument. One is to accuse the debater of "betraying our moral tradition", or even of being immoral, just because he dares questioning some received moral principle. Another tactic is to resort to prudential considerations — e.g. "What you propose to do is morally right, but I won't go along because it is too dangerous". Finally, one may claim that there can be no rational moral arguments (only shoutings and countershoutings) because morality is a matter of emotion or of interest not of reason. The first two tactics are of particular interest to moral philosophy for being morally wrong. This is evident in the second case, a bit less so in the first. Those who resort to the first tactics may do so in clear conscience and just out of ignorance concerning the changeability of moral principles. By being dogmatic they are also being, albeit unwittingly, intellectually dishonest. Honesty cannot be fully exerted without some critical disposition.

In short, reason is as necessary in moral and ethical matters as anywhere else. In particular, explicit logical analysis is necessary to validate or invalidate moral arguments and to evaluate ethical doctrines. And such analysis only calls for ordinary logic.

1.2 *Ethics and Epistemology*

If the good and the right are always totally subjective or conventional, then their corresponding theories, namely axiology and ethics respectively, could be independent of epistemology or the theory of knowledge. But if deliberation about values and morals are at least sometimes guided by factual knowledge, then axiology and ethics must involve epistemology.

Now, rational people, whether craftsmen or technologists, lawyers or physicians, scientists or scholars, usually give reasons, good or bad, for some of their valuations and decisions, in particular those of the moral kind: When questioned they do not resort to taste or custom but to facts. For instance, we say that object A is good for person B as shown by experience C; or that action A is right for individual B because it meets a basic need *C* of *B*. So, as a matter of fact, far from *opposing* values to facts, rational people often evaluate and decide in the light of their knowledge of facts: They seek knowledge in order to evaluate correctly or act rightly. Hence valuation and deliberation about action, as well as decision and action, *connect* values to facts instead of detaching them. Consequently axiology and ethics, far from being alien to epistemology, must somehow be related to it.

Irrationalists and conventionalists are bound to countenance absurd valuations, decisions, and actions. (Curiously enough, both positivists and existentialists believe in the basic irrationality of all valuation and decision: See Albert 1985.) Subjectivists, if consistent, can easily skirt moral problems, for they have no reason to believe that other people have the same basic needs, hence the same basic rights and duties, as they. An anti-realistic epistemology encourages one to overlook the differences between what is objectively good or right, and what only seems to be good or right, since such an epistemology has no way of telling reality from appearance. On the other hand a realistic epistemology helps us make correct evaluations and decisions, and perform right actions, for realism enjoins us to face facts instead of telling us that everything concerning values and morals is subjective, whence "anything goes" in their regard.

A realistic epistemology will encourage the testing and eventual review of moral codes and ethical theories, for it will ask how they fare in the real world. Now, any test of ideas is both conceptual and empirical. (See Vol. 6, Ch. 12.) The *conceptual test* of a moral code consists in checking whether it matches the underlying value system, i.e. whether the rules in the code enjoin us to realize our values. And an *empirical test* of such a code consists in checking whether the observance of the code does indeed bring about the desired results. In principle any such test could be experimental, but in practice it is usually observational.

As for the test of ethical theories, we must begin by recalling that, in our view, an ethical theory is a theory about morals, in particular about the roots and functions of moral codes. Such a theory may or may not fit the facts in a satisfactory fashion until new order. For example, hedonism is false because (*a*) healthy people are naturally active and the exclusive pursuit of pleasure causes boredom, and (*b*) utter selfishness is discouraged by all societies, even those the official ideology of which is egoistic. Likewise sociologism, or the view that morality is nothing but a social control device, is false for overlooking moral sentiments as well as the inner springs of action, i.e. needs and aspirations. By resorting to some findings of psychology and social science we have combined conceptual with empirical test procedures. At the same time we have taken it for granted that ethics is or can become scientific, whence there is no frontier between ethics and science. (See Bunge 1960, 1961. More on the evaluation of ethical theories in Pollock 1988.)

However, it might be objected that morality and ethics have nothing to do with epistemology and science because moral codes prescribe actions, and ethical theories justify morals, whereas epistemology and science describe and explain. In other words, it would seem that, whereas science and epistemology are about what *is*, morals and ethics are about what *ought to be*. But this objection is incorrect for the following reasons. Firstly, every science contains prescriptions or rules, e.g. for making observations and experiments, and even for stating laws; and epistemology has a normative chapter, namely methodology, which studies such prescriptions. Secondly, any realistic moral code is supposed to fit real people, with their needs and wants, abilities and shortcomings, rather than ghosts or saints. Unrealistic morals are useless, except perhaps as subjects of doctoral dissertations. Thirdly, putting any moral precept into practice amounts to transforming what *is* into what *ought to be*, i.e. to guiding the system of interest (which may be the agent herself) from its current state to the desired or goal state. And if we wish to accomplish such a transformation in an efficient manner, we had better get hold of the relevant scientific or technological knowledge.

Another reason for clarifying and strengthening the ties between the theory of morals and the theory of knowledge is this. Moral action presupposes moral responsibility, which in turn requires an adequate understanding of the problems involved as well as of the available means. The philosopher does not help attain such an understanding if he holds the conventionalist thesis that decision is prior to knowledge rather than the other way round. A realist epistemologist, on the other hand, will not fail to see that responsibility involves knowledge: that an irresponsible person is one who does not know better, or who ought to know better. Responsible individuals do not rush heedlessly into action but make use of the knowledge required to bring about the desired outcome: they join the moral and the prudent. (Recall Ch. 6, Sect. 2.2.)

The Stoics were right in holding that there is no virtue without knowledge. Moreover we may define "wisdom" as virtue together with knowledge.

However, there are cases where information may have to be partially suppressed in order to guarantee moral or legal impartiality. For example, justice is represented as a blindfolded woman, to suggest that we should all be equal before the law. Rawls's "veil of ignorance", to be worn by the distributors of goods and burdens in the mythical "original position", is an application of the same idea to distributive justice. School uniforms have a similar function, namely to veil differences in "initial endowments". The refraining from telling certain disagreeable truths out of compassion is a further example of deliberate suppression of information. So is the avoidance of information overload in scientific or scholarly research, but this time in the service of insight. Refraining from using information of some kind for the sake of impartiality, compassion, or insight, is very different from making it up, as is the case with the "useful lies" of the pragmatists and with the pretenses or *as ifs* of the fictionalists. Hence none of the cases cited above, which refer to the utilization of knowledge, not to its gathering, supports an irrationalist, fictionalist, or pragmatist epistemology.

In short, realistic, hence efficient, morals, and true theories about moralities, call for the assistance of a realistic epistemology.

1.3 *Ethics and Ontology*

Axiological and ethical absolutism hold that values and morals are above mundane matters, whence axiology and ethics have no ontological commitment: like logic, they hold in all possible worlds and come what may. Thus Tolstoy's Kantian Levin in *Anna Karenina* (Pt. 8, Ch. 12): "If goodness has a cause, it is no longer goodness; if it has consequences — a reward — it is not goodness either. So goodness is outside the chain of cause and effect".

In our view every factual discipline has ontological presuppositions and in turn poses ontological problems. (See Vol. 3, Introduction and Vol. 6, Ch. 14, Sect. 2.1.) Since value theory and ethics deal ultimately with facts, namely valuations, decisions, and actions, they must depend on ontology. Take for example the old question: What is worse, a natural calamity such as the collapse of the solar system, or a sin, however venial? The answer depends on one's ontology. According to a supernaturalistic ontology, since nature is far below the City of God, stealing a penny is worse than the destruction of the ozone layer, whereas according to naturalistic ontology it is the other way round.

We take it that ethics presupposes a solution to certain ontological problems, such as those of the nature of mind and the possibility of freedom. In turn it poses certain ontological questions, such as whether materialism is compatible with high moral ideals or whether, on the contrary, it is necessarily piggish. Let us take a quick look at some of these problems, starting with that of the nature of mind.

Consider two extreme positions: those of eliminative materialism and of spiritualism. According to the former we do not have minds but are genetically programmed automata. Clearly, if this is true then we can have neither moral feelings nor a conscience, and consequently morality makes no sense except perhaps as a reward and punishment schedule designed exclusively to protect our precious genomes. On the other hand, if we have an immaterial and immortal soul, we may feel free to use the body: our own with dissoluteness, other people's with cruelty. After all, in either case we shall be affecting "only" the material component of the person, which is perishable anyway. So, both vulgar materialism (or physicalism) and psychoneural dualism condone incontinence and cruelty. (But of course consistency is rare. Thus Epicurus, a physicalist, was an ascetic and taught not to harm; and Mother Teresa, another ascetic, alleviates the sufferings of the dying.)

On the other hand, according to emergent materialism (Vol. 4 and Bunge 1981), incontinence and cruelty are wrong, because whatever is done to a person's body affects his mental processes, which are brain processes of a very special kind. The physicalist hangman feels no compunction because there is no mind to be murdered; not does the dualist hangman, because the soul is detachable from the body. Only the emergent materialist has a good case against the death penalty and, a fortiori, against mass murder, in particular war. He knows that every person is the unique, unrepeatable outcome of an original genetic combination and a unique learning experience; that every normal person has a mental life, in particular a pricking conscience, and can make a unique contribution to other people's lives; and that he has only one life, so he had better make the most of it for himself and others.

Likewise the emergent materialist admits free will and even attempts to explain it in terms of spontaneous (stimulus-independent) neuronal activities (Vol. 4, Ch. 4, Sect. 5.1, Hebb 1974 and Bunge & Ardila 1987). Thus he rejects both classical determinism ("Freedom is illusory") and classical indeterminism ("Freedom is for real and it violates the laws of nature"). In particular, he rejects the sophism "Freedom is the knowledge of necessity", coined by Spinoza and adopted by Hegel, Comte, and Engels. Curiously, this was the secular version of the Christian dogma that only the knowledge of the true religion can make us free. In either case freedom is equated with the awareness of servitude. But of course this is the very antithesis of active freedom or *potestas agendi*: in a world that is preordained either naturally or preternaturally there is no choice, let alone the possibility of creating new alternatives to whatever options are given us.

This is not to deny that knowledge (of chance as well as of necessity) facilitates the exercise of freedom. In particular, knowledge of some laws of physics and chemistry has allowed engineers to design artifacts that have freed us *from* some natural constraints, others that enable us to *do* things that our ancestors did not even dream of — and still others which have created new servitudes. Therefore not only determinists like Spinoza and Voltaire but also indeterminists like Eccles and Robinson (1985) are wrong in holding that free will is incompatible with the laws of physics and chemistry.

In short, freedom is neither the knowledge of necessity nor an attribute of the immaterial mind: instead, it is the ability of some highly evolved brains to make choices and create new alternatives overcoming, though not overlooking, certain external constraints. Without freedom there would be neither virtue nor sin, neither credit nor blame, neither pride nor remorse. It makes no sense to expect a praying mantis to feel remorse for eating her mate, or to blame a robot soldier for killing, because neither of them has the ability to change its program. Only humans, and perhaps a few other higher animals as well, can be credited for goodness or blamed for evil. Emergent materialism makes sense of morals because it accounts for freedom.

However, let us not exaggerate the range of our freedoms. We are never totally free, self-directed, or autonomous, only partially so, for we cannot escape our animal condition or our social bonds. Even the freest of individuals satisfies certain biological and social laws and constraints. Moreover, the great majority of humans are shackled to misery and oppression. Those few who do enjoy a certain measure of freedom do so for observing at least some of the rules of the game, and often at the price of the servitude of others, particularly their spouses and underlings.

Given that we are only partially autonomous (self-directed), any moral philosophy that postulates either that man is autonomous (e.g. Kant's) or heteronomous (e.g. Spinoza's) is unrealistic and therefore useless. We are free to choose between right and wrong, or good and evil, provided we can tell them apart and have the courage and the means to do so. Even so, our freedom to grapple with moral problems is rather limited. Firstly, most moral problems are thrust upon us, often unexpectedly, so that we may not be well prepared to solve them: We do not enjoy the freedom of inventing moral problems the way we invent chess puzzles or mathematical problems. Secondly, the moral problems we confront arise in social circumstances that are for the most part not our own creation and are largely beyond our control. Thirdly, whether we seek to solve such problems on impulse or after careful deliberation, we always have limited resources (time, strength, connections, etc.) to solve them efficiently. In short, we are free but only in some regards and to some extent.

A third ontological problem that must be solved before engaging in ethical theorizing is that of the reality of conflict and harmony, in particular competition and cooperation. An ethical theory that presupposes the Heraclitean view that all is conflict, or Leibniz's view of universal harmony, is too unrealistic to succeed as a guide to life. An effective ethical system must be based on an ontology admitting both competition and cooperation, and even combinations of both, on all levels of reality. Ours does so (Vol. 4, Ch. 4, Sect. 5.4, and Appendix B, Sects. 2.3 and 2.4, and Bunge 1981.) Hence it makes it possible to combine egoism with altruism, self-reliance with mutual help, competition with cooperation. We shall come back to this in Ch. 10, Sect. 2.4.

So much for some of the ontological presuppositions of ethics. Now for the problem whether materialism leaves any room for morals. We saw above, when discussing the problem of free will, that it does. Yet since Antiquity spiritualists have been claiming that materialism is immoral, because high ideals are the opposite of base matter. Actually spiritualism, not materialism, is the enemy of moral idealism for holding that all of the moral principles are immutable and lack a natural and a social basis, so that they must be accepted as supraempirical dogmas for all time (Ingenieros 1917). However, it is true that vulgar materialism or physicalism, particularly when combined with classical determinism, has just as little use for morals as fatalism, whether secular or religious. But, as we saw a moment ago, emergent materialism is very different, for it admits both (bounded) freedom and the need for a morality rooted to individual and social welfare.

In short, ethics and ontology are intimately connected. Ontology is ethics-free, but ethics has ontological inputs (presuppositions) and outputs (problems). Hence a false ontology may inspire a wrong ethics, and this may in turn raise ontological pseudoproblems. Some of the many woes of contemporary moral philosophy have ontological sources. But this will not be seen unless it is realized that all of the branches of philosophy hang together. Recall Figure 0.1 (Introduction).

1.4 *Ethics and Science*

Ethics and science are related in two ways: There is the ethics of scientific research, and there is the science of ethics, or scientific ethics. The former is the investigation of the moral code governing scientific research, whereas the science of ethics is the scientific study of the roots and functions of moral codes. We have attempted to sketch the latter in the foregoing, so here we shall only grapple with the former.

Some of the rules that govern scientific research are strictly moral, whereas others are methodological as well as moral, and finally others are only methodological. For example, the rule that commands giving credit where credit is due, is moral but not methodological. Indeed, a plagiarist who steals good results is a baddie but not necessarily an incompetent worker. On the other hand the rules that command not to doctor empirical data, much less to make them up, and to seek and accept expert criticism, are both moral and methodological. The same applies to the rules that enjoin us to seek and defend the truth, and refrain from covering up the defects or limitations of one's research, withholding information, and distorting science in the interest of ideology. Finally, the statistical sampling rules are purely methodological. But of course any deliberate disregard for methodological vouchsafes is a breach of the ethos of science. See Table 8.1.

We call *endomorality* the moral code of scientific research. Unless observed, travesty is bound to result. Therefore any definition of "science" which overlooks the endomorality of scientific research is defective, possibly to the extent of being incapable of helping diagnose fraud. In other words, any adequate definition of the concept of science must include a reference to its moral code, which is designed to encourage and protect the search for truth, i.e. the pursuit of adequate (though not necessarily iconic) models of reality. (For more on the

TABLE 8.1. The partial overlapping of scientific method and scientific morals: a sample.

Moral rules	Give proper credit. Do not make experimental animals suffer. Do not waste.
Moral and methodological rules	Do not doctor data. Check for truth. Seek expert criticism.
Methodological rules	Begin by choosing a problem and formulating it clearly. Look for systematic errors in the experimental design. Disregard data involving discrepancies larger than 3σ.

morals of science see Cajal 1941, Bronowski 1959, Snow 1959, Bunge 1960, Sjoberg Ed. 1967, Monod 1971, Max-Planck-Gesellschaft 1984.)

The endomorality of science harmonizes with a realistic epistemology and it clashes with conventionalism, constructivism, fictionalism, and pragmatism — all of which deny the possibility of objectivity. These anti-realistic epistemologies do not account for the aim of scientific research, which is to understand reality, and consequently they do not fit the ethos of science, which is designed to encourage and protect the search for objective truth.

Science is *the* field where success depends exclusively on the finding of *true* data, the construction of (approximately) *true* hypotheses, and the design and utilization of efficient methods for finding *true* data or for checking hypotheses for *truth*. The Nobel prizes in natural science are supposed to be earned for outstanding true findings concerning reality, or for breakthroughs in the search for new methods for finding factual truths — not for introducing conventions, inventing fictions, or designing commercially or politically successful contraptions. On the other hand in technology success depends on social demand as well as on truth. And in business and in politics falsity often pays off, and only failure is penalized. Therefore the pragmatist redefinition of "truth" as that which promotes life and success, and of "error" as that which does not, is orthogonal to the endomorality of science. (See a vehement attack on the pragmatist doctrine of vital lies in Lee 1912.) The current popularity of antirealistic epistemologies is an alarming indicator of the popular misunderstanding of science.

We call *exomorality* of a field of knowledge that part of its moral code which concerns the social responsibility of its practitioners as such. And we submit that the exomorality of the community of basic scientists is composed of a single injunction, namely *Thou shalt produce and teach good science.* Basic scientists are to blame if they fail to honor this precept, in particular if they commit frauds or teach lies. But they cannot be blamed for unwittingly producing knowledge that others may use for evil purposes. Indeed, given that the goal of basic scientific research is to understand the world, not to control or alter it, the basic scientist as such has no opportunity for evil-doing — and usually no ability either. The nuclear bomb was designed by engineers not by nuclear physicists: the latter only provided some of the knowledge utilized by the former. Chemical and bacteriological warfare are analogous. In short, basic science is morally (and politically) neutral with regard to society, in that its sole function is to enrich culture (Bunge 1988a).

This is not to say that scientists are incapable of influencing the social process. They do so every time they alter our world view, and every time they find something that is utilized by industry or government to manage or mismanage the world. Suffice it to recall the profound and lasting influence exerted by the scientific work of Newton and Rutherford, Darwin and Mendel, Smith and Marx, to mention only a few. All original scientists are unwitting actors, or at least extras, in the human tragicomedy, because science happens to be the nucleus of modern culture and the wellspring of technology, which is in turn the engine of modern industry. Still, it remains true that the social responsibility of scientists is not nearly as heavy as that of technologists, businessmen, bureaucrats, and politicians.

Still, it is often claimed that, whereas natural science may well be morally neutral, social science cannot because it deals with social problems, many of which involve social conflicts, e.g. between the propertied and the dispossessed, so that the social scientist must take sides with one of the conflicting parties. This is certainly the case with applied social science and sociotechnology, for they are concerned with social policies and programs. But it is not the case with basic or pure social science, which, if partisan, is ideology not science, hence cannot serve as an objective basis for the design of social policies and programs. (Vol. 7, Ch. 4, Sect. 4.1, and Ch. 5, Sect. 4.2.)

Mainstream economics is a case in point. No doubt, some theories in this field are suspect for assuming a purely economic model of man (see e.g. Sen 1977 p. 336) and for overlooking the emergent properties of social systems. Worse, some of them are undisguised manifestos in defense of special interests groups: see e.g. Robinson & Eatwell (1973). In particular, the neoclassical theories not only assume but also recommend selfishness, and they ignore such problems as poverty, militarism, and environmental degradation: they are not only scientifically defective but also immoral. On the other hand there are a number of perfectly objective and morally neutral ideas in economics, that can be used now for good purposes now for evil ones. Thus a Leontief input-output matrix, an econometric model, and a model for the rational exploitation of a fishery or the rational management of a firm need not be ideologically loaded, hence morally committed.

In short, basic science, whether natural or social, has a rigorous internal moral code. Its only social responsibility is to keep healthy.

1.5 *Ethics and Technology*

We shall understand 'technology' in the broad sense, namely as the science-based discipline engaged in designing (not manufacturing) artificial things and processes, whether inanimate like a computer, living like a cultivated plant, or social like a factory. (See Vol. 7, Ch. 5.) Accordingly not only engineering but also agronomy, management science and the law are technologies.

Technology has its own endomorality or internal code of conduct. It overlaps partially with that of science because both seek the truth, hence both call for intellectual honesty. But for technology truth is a means to an end, namely usefulness to somebody; hence accuracy and depth are expendable if they conflict with practicality, cost efficiency, or some other typically technological desideratum. Indeed, the technological imperative is "Thou shalt only design useful artifacts or processes, even at the price of some inaccuracy or some loss of depth". This stands in sharp contrast with science, where accuracy and depth are paramount whereas practical utility plays no role. (See Vol. 6, Ch. 14, Sect. 2.2.)

The endomorality of technology is more lax than that of science. For one thing whereas the scientific community penalizes or even ostracizes the plagiarist, technologists are often encouraged to copy or steal. In fact, many technologists are asked to imitate existing artificial things or processes while marginally observing the patent law, and some specialize in industrial espionage. None of these activities reflects on their professional standing. As well, whenever high quality conflicts with low cost or large benefit, the latter is likely to get the upper hand. See Table 8.2.

Whereas the exomorality of the scientific community is tiny ("Supply good science"), that of technology is or ought to be huge because the implementation of any technological project alters reality, and most artificial changes are likely to benefit or harm certain groups of people. Just think of the social and environmental changes caused by technological innovations over the last two centuries. In view of the enormous power that technology gives business and government, the technological community ought to adopt a far more strict code of conduct including in the first place the norm "The survival and welfare of humankind should take precedence over the interests of employer or client". This rule entails (informally to be sure) the norms "Minimize the undesirable social and environmental impact likely to be produced by any technological innovation", and "Refuse to obey, or even blow the whistle, when ordered to design anything that is likely to hurt the public interest". Regrettably, most technologists are more eager to embark on new exciting projects, or to humor their employers or clients, than to allow their hands to be tied by moral precepts. This attitude is exemplified by E. Teller's comment on J. R. Oppenheimer's objections to constructing a hydrogen bomb: "He objected to the H bomb merely

TABLE 8.2. Some characteristic values and moral norms of science and of technology.

Values	Moral norms
Basic science	
Truth	Do not lie.
Depth	Do not withhold information.
Research freedom	Do not skirt problems.
Originality	Challenge received opinion if necessary.
Technology	
Usefulness	Do not design anything useless.
Truth	Be truthful to employer or client.
Efficiency, reliability	Be loyal to employer or client.
Low cost and low risk	Check for flaws that may put lives or monies at risk.

on moral grounds". The transition from "morals" to "mere morals" accompanies the metamorphosis of the moral agent into a moral moron.

However, in most cases the engineer working in civilian industry or in government does not often face any moral conflicts and, when these do come up, he can solve them in consultation with his employer or client. Here is a short list of types of problem of this kind. (More in Schaub & Pavlovic Eds. 1983.)

(i) *Quality control* in design and production. Economic motivations: uniformity and business continuity. Moral motivation: honesty. (Limitation: Quality and cost are inversely related, whence a compromise between the two must be worked out.)

(ii) *Human engineering* in the work place. Economic motivations: cost minimization and worker satisfaction. Moral motivation: personnel welfare.

(iii) *Risk control*: keep risk to personnel, consumer and public below the tolerance level. Economic motivation: high risk products or processes jeopardize prestige and give rise to expensive law suits. Moral motivation: protection of well-being.

(iv) *Social impact*: minimization of undesirable social changes. Economic motivation: abrupt social disruption leads to economic disturbance and labor unrest. Moral motivation: minimization of sufferings caused by the need to adapt to quick drastic changes.

(v) *Environmental impact*: minimization of undesirable environmental changes. Long term economic motivation: protection of natural resources. Moral motivation: respect for public property (air, water, natural beauty, etc.).

In all of the above cases the long-term interests of the technologist and his employer coincide with those of the workers and the public, so that the moral problem can be given a technological solution. But in other cases the technologist comes up against economic greed, political ambition, or sheer ignorance. In such cases he has a moral obligation to blow the whistle, i.e. of objecting, and even going public, if pushed into projects that are technically flawed or likely to hurt the public interest (Nader, Petkas & Blackwell Eds. 1972).

Apparently few technologists ever feel moral qualms, and even fewer bring moral problems to the public attention. And when they do so they can expect to be fired (Westin 1986). In a few places there are laws protecting the whistle-blowers from reprisals, and professional societies may be expected to lend them lukewarm support. But the only effective way of preventing the implementation of projects likely to endanger public welfare is active citizen participation. This can go from public protest and public consultation (e.g. through referenda) to interactive planning involving representatives of governmental and non-governmental organizations in the planning process itself. We shall return to this subject in Ch. 10, Sect. 2.5.

The technologists whose work is bound to have a large and immediate impact on society are the sociotechnologists, in particular the normative economists, jurists, political scientists and policy analysts acting as government employees or consultants. They all deal with policies and large scale plans, hence they all face moral problems, because every social policy and every social program is likely to affect the well-being of many people — or it would not be designed. (Vol. 7, Ch. 5, Sect. 4.2.) Inevitably, the sociotechnologist is often torn between the wishes of his employer or client, and the demand to tell the truth. When such conflict involves the public interest, it is his professional and moral duty to advocate the best policy and criticize its rivals (Bodde 1986).

It is often required that any social policy or program satisfy Pareto optimality: that it favor at least one member of society without harming others. And harm is considered admissible if the victims can be compensated for — the Kaldor-Hicks criterion. The two criteria are exclusively economic: they presuppose that everything can be priced. In particular, they overlook economically imponderable damages such as those caused to environment or culture. Moreover Pareto optimality is not really optimal. It is best to aim for large scale actions from which *everyone* stands to gain, such as massive job creation, sanitation, or literacy programs. Since the economy is only one of the subsystems of a society, asking economists to design all social policies and programs is to court disaster and to abdicate moral responsibility.

The question is often asked whether technological progress is good or bad. There are three possible answers to this question: technolatry, technoclasm, and technorealism. *Technolatry* is of course the naive view that all technological progress is good. This view is false because (*a*) many technological advances occur in the field of military technology, which is not surprising since more than half of the engineers in the world are employed in that field, and (*b*) much technological progress makes us increasingly dependent upon vulnerable or intensive energy-consuming and entropy-increasing systems or processes. (See Neirynck 1986.)

Technoclasm, the thesis that all technological advance is bad, is equally wrong, if only because (*a*) technological progress has doubled the life expectancy, and halved the working hours, in the industrialized countries in the course of two centuries, and (*b*) we need technology to repair the ravages caused by unbridled industrialization, to tap alternative energy resources, and to ensure the survival of civilization — only, the technology we need is strictly civilian and dedicated to the welfare of humankind.

The rational alternative to both technolatry and technoclasm is *technorealism*, which may be summarized as follows. Firstly, whereas some artifacts are univalent, others are ambivalent. Among the former some (e.g. tooth brushes) are good in certain respects, whereas others (e.g. offensive weapons) are bad in all respects; as for ambivalent artifacts, they are good or bad depending on the jobs they are used for — e.g. a knife can be used for cutting bread or throats. Secondly, technology is a means not an end; hence there can be either progress or regress through technological advancement. Thirdly, all technological progress aiming at enhancing human welfare is good in some respects and bad in others — e.g. in its unintended negative side effects. Fourthly, all advances in offensive military technology are bad, and so are all of the sociotechnological advances aiming at exploiting or oppressing. Fifthly, every technological innovation must be assessed in its social context: we must ask whether society really needs it and can afford it. Sixthly, every large scale technological project should be approved or rejected by the citizenry on the advice of experts. (See Agassi 1985.) We shall come back to the problem of the democratic control of technology in Ch. 10, Sect. 1.4.

Finally, let us tackle the question whether science and technology can only determine means, never goals, for the latter are always determined by needs or wants. There is some truth in this view. In fact, technology helps meet any need and any realistic desire. But it is not true that science and technology have no business in the choice of goals. Firstly, the means-goal distinction is not fixed but fluid: What begins as a means may end up by becoming a goal and conversely — e.g. killing in self-defense may turn into killing for pleasure. (See Schlick 1925 Ch. 13.) Given such interconvertibility of means and goals, science and technology may help determine goals, albeit indirectly. Secondly, science and technology alone can help us evaluate goals prudentially, by telling us whether a given desideratum is feasible, who can expect to benefit from it and who be harmed by it, and so on. We conclude then that science and technology can help determine goals as well as means. But we hasten to add that science and technology are not to replace morality: they can only supplement it. Knowledge without moral conscience may only be the tool of wickedness.

To sum up, whereas in the worst of cases scientific knowledge is worthless, technological knowledge can be harmful, hence it must be subjected to external regulation. After all there is such a thing as forbidden knowledge (Rescher 1987).

1.6 *Summary*

Far from being an autonomous discipline, ethics depends not only upon axiology but also upon logic, epistemology, and ontology. Indeed, there is no correct moral argument or ethical theorizing without adherence to logic. Nor can we make morally right decisions, or set up and check ethical principles, unless we assume a realistic epistemology. Only the moral nihilist can adopt an anti-realist epistemology, and only the moral conservative can fail to understand that we must get to know the world before we attempt to foist our moral principles upon others. What ought to be cannot be unless we learn what it takes to exercise our rights and perform our duties. Nor could we be efficient moral agents in a ghostly world, for all action is a lawful transformation of a concrete thing by another — the agent. Hence any realistic ethical theory presupposes an ontology of concrete (material) things behaving in a lawful fashion. Don Quixote was inefficient because he mistook his own fantasy for reality. To be sure his page Sancho was enough of a realist and a materialist — but, alas, he had little moral sense.

Ethics also has close ties with science and technology. To begin with, ethics should and can become scientific, i.e. consistent with the scientific findings about our needs and desires, sentiments and social condition. In addition, the moral philosopher may wish to study the moral norms that govern or ought to govern the specific activity of scientists and technologists. He cannot fail to notice that, unlike basic science, technology has a heavy social responsibility, for every major technological project has moral (or immoral) presuppositions and implications. If he does pay attention to the moral dimension of technology, he cannot remain indifferent to the ways technology is being misused: he will join in whistle blowing.

In sum, ethics does not exist in a vacuum. Not only does it grapple with morals, hence with values. If scientific and relevant it uses findings

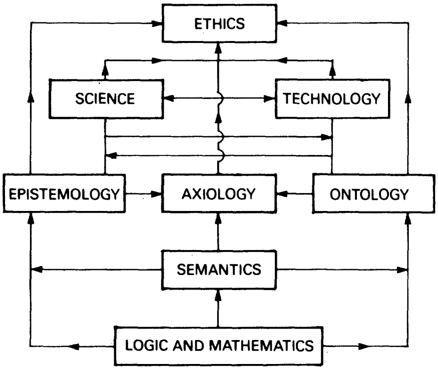


Fig. 8.1. The fields of knowledge presupposed by scientific ethics.

of other philosophical disciplines as well as of science and technology: See Figure 8.1. Isolation breeds irrelevance and obsolescence.

2. ETHICS AND ACTION

2.1 *Praxis and Ethics: Generalities*

In this and the subsequent sections we shall be concerned with some of the moral problems that arise in the exercise or practice of a skill or a branch of knowledge concerned with the alteration of real things. We shall not deal with the design of artificial things, such as new drugs, or of processes, such as medical treatments, but with the implementation of such designs. Such implementation goes through four phases: prescription, supervision, actual execution, and evaluation. For example, a physician may prescribe a medical treatment, whereas nurses, laboratory technicians, physical therapists and others will help apply and supervise the treatment, and the doctor herself or some colleague will evaluate its results. We shall focus here on professional practice rather than on the original conception and laboratory test of the design in question. We shall concentrate on the moral aspect of actions as specific and yet so similar as those of the general practitioner and the appliances serviceman, the lawyer and the paramedic, the civil servant and the teacher, the manager and the foreman, all of whom see to it that the state of certain systems be changed in certain prescribed manners.

Every professional code is a set of rules, some technical, others moral, and still others mixed. The technical rules prescribe in broad lines how to proceed, the moral ones what to do in cases of moral conflicts — e.g. between means and ends. Though distinct, the technical and the moral rules are mutually related. A first link is between professional competence and professional morality. The rule states that professional competence is necessary, though insufficient, for professional morality. (Shorter: $M \Rightarrow C$.) Put negatively: It is morally wrong for people lacking professional competence (as a result of lack or deficiency of professional training) to pursue a professional activity, for they can cause harm. For example, people who have not trained as scientific psychologists are not morally qualified to tamper with mental patients.

A second link is between professional conduct and the interests of the person who employs the services of the professional. The rule is that the professional shall use his best professional knowledge and judgment, and is to work hard, to advance the interests of his employer, client, or patient. A third rule is that, in case of conflict between the professional's employer, client, or patient, and the public interest, the latter is to prevail.

The teacher's professional code is one of the simplest, for it boils down to three rules: "Learn well the material you are to teach", "Do your best to help pupils learn and enjoy the learning", and "In case of conflict between truth or didactic efficiency, on the one hand, and authority on the other, bow to the former". All three rules may be regarded as moral as well as technical. The hardest to observe is the third: witness the uncounted Monkey Trials since the beginning of formal education. By preserving academic freedom, the institutions of higher learning have minimized, though not eliminated, the conflicts between truth and authority. Still, almost anywhere in the world it is imprudent to teach certain philosophical, political, or even biological theories.

It takes moral courage to practice any profession. The seriousness of the moral predicament of a professional is roughly proportional to his ability to protect or harm other people's interests — vital, economic, political, or cultural. Thus political and business ethics are trickier than medical or legal ethics, which in turn are harder than the ethics of craftsmen. Let us take a quick peek at a few special professional ethics.

2.2 *Bioethics*

The main moral problems involved in biomedical research and practice are those raised by patient treatment, tampering with life, human and animal experimentation, and genetic engineering. The first in the list is also the oldest, and it is still being dealt with, in the main, on the basis of the Hippocratic Oath, the central precept of which is "Do no harm". Medical progress is making this precept increasingly hard to apply, for a number of medical treatments involve high risk, high cost, or both. A physician who is to adhere strictly to the Oath would refrain from recommending brain or open heart surgery and from prescribing drugs with serious side effects, and he would definitely refuse to practice euthanasia.

It seems then that the negative utilitarian precept "Do no harm", though appropriate at a time when most medical interventions were ineffectual or counterproductive, has become obsolete. (Recall Ch. 7, Sect. 2.5.) Medical ethics and the criminal law have to be reformed on the basis of the more general maxim "Enjoy life and help live". If a patient cannot enjoy life for being afflicted by a severe illness, his doctor should be free to offer him any promising treatment, regardless of the risk it may involve. And if all hope were lost, the patient should have the right to be put to painless death. Physicians have the duty to help live lives worth living, not to prolong hopeless misery. The law should accordingly be reformed in such a way that physicians be guided solely by science and compassion, not by fear of legal sanctions. This reform is particularly necessary wherever the insurance industry is ready to compensate patients or their dependents for any real or imaginary medical malpractices.

Clinical psychology and psychiatry pose particularly interesting ethical problems. The very first problem is that of professional competence. Many psychotechnologists are not properly qualified to tamper with the mind, for not knowing any scientific psychology: they are dangerous quacks. A second problem is that many people who are willing to take doctors's orders object to a clinical psychologist's using behavior therapy to correct some behavioral, emotional or cognitive disorder. This attitude has a theologico-philosophical root, namely mind-body dualism. For dualists it may be all right to tamper with the material body, but sacrilegious to manipulate the immaterial and possibly immortal soul. The biological view of mind as a brain function (Vol. 4, Ch. 4) avoids this groundless myth and suggests that a psychological or psychiatric treatment is ethical provided only that it is (*a*) scientifically sound, (*b*) practically effective, and (*c*) compassionate.

The more serious moral problem raised by psychotechnology appears when the patient engages systematically in serious antisocial behavior, such as rape or murder, and refuses to submit to treatment. Extreme libertarians claim that nobody has the right to force such patients to undergo any treatment aimed at correcting their misbehavior. Hence libertarians are left with only two options, neither of which is in the best interest of either the subject or society, namely life sentence and capital punishment. In our view, if a psychiatric or psychosurgical treatment can transform a professional criminal into a normal individual capable of being reinserted into society, such treatment should be forced on him for his own good as well as for the good of society. (More on the so-called paternalistic society in Sect. 2.3.) Again, this course of action fits in with the maxim "Enjoy life and help live".

We discussed earlier (Ch. 6, Sect. 1.2) the moral problems of tampering with life, chiefly in relation with abortion and euthanasia. Let the following suffice here. The so-called pro-life advocates, who attack abortion but not mass-murder (i.e. war), argue that an embryo and a fetus are human beings, presumably for having being endowed with an immaterial and immortal soul. But the point is that neither embryos nor fetuses satisfy any of the accepted definitions of a *person*: both are *parts* of a woman and neither has a *mental* life. Indeed (*a*) a fetus cannot become a separate individual unless it has been delivered, and (*b*) the mental faculties emerge only after birth and they take several years to develop. So, the pro-life movement has no scientific basis. Furthermore it violates the norm "Enjoy life and help live", for (*a*) it interferes with freedom and well-being, and (*b*) it helps the multiplication of unwanted children, most of whom are bound to be unhappy and a burden to society. Abortion is the lesser of two evils. It is an evil that can be avoided by using contraceptives and discouraging sexual promiscuity. A pro-love movement would be more moral and constructive than the pro-life movement.

As for euthanasia, the very first thing to do about it is to forget the etymology of the word, i.e. 'good death'. It is a poor defense of euthanasia to claim that certain people are "better off" dead than alive, or that they maximize their own utility by being put mercifully to death: dead people are not people, hence they cannot have any utilities; they cannot even be "worse off" than the living. The question is not one of good death but of bad life. If life can no longer be enjoyed, e.g. because of extreme pain or helplessness, there is no point in prolonging it. The patient who asks for a lethal injection does not compare his present state with that of his own corpse and decides that the latter is in a better state: he simply cannot stand life any longer. To deny his last wish is sheer cruelty. (More on euthanasia, abortion and related matters in Spicker & Engelhardt Eds. 1977, Ferrater-Mora & Cohn 1981, and Cragg Ed. 1983.)

A minor way of tampering with life is to use medicine to improve the performance of athletes. The unethical medical coach of a commercial sports team may be asked to administer massive doses of masculine hormones in order to build herculean muscles, or amphetamines to allay fatigue, disguise pain, and create rage. The most effective way of eradicating such malpractices is to discourage professional sports, which serve the same functions as gladiatorial games and bull-fighting, namely to entertain the uneducated, to raise their level of violence, and to divert their attention from social issues. A good society invests in amateur sports and in general health care what our societies spend on commercial sports and Olympic games.

Let us now examine experimentation on humans and other animals. Over the past two decades the international biomedical community has reached a consensus concerning the conditions under which medical experiments may be conducted on human subjects. These conditions include informed consent, professional competence, correct experimental design, supervision by a professional committee, and compliance with the law of the land. Still, as with any other norm, there are areas of penumbra, hence sources of controversy. For example, should an experiment be performed even if its underlying hypotheses contradict a solid body of biomedical knowledge or if, far from involving any hypothesis, it is conducted merely "to see what happens", in adherence to the classical empiricist methodology? (Answer: No. In the first case because it would be immoral, in the second because it would be scientifically barren on top of being immoral.) How are the interests of children, the physically or mentally handicapped, and prison inmates to be protected? (Answer: By the conditions listed above, except that the responsibility of the supervising committee is now greater.) In every case the rights to life and well-being should be protected, and whenever risk is involved the experimental subject should be compensated. For example, a life prisoner should be allowed to gamble his life for freedom. But it would be morally monstrous to experiment on prison inmates without their consent and without compensation, for this would not help them live. Again, the standard is “Enjoy life and help live”.

How about biomedical experiments on animals? Many societies for the prevention of cruelty to animals, and more recently a movement for animal rights, have opposed animal experimentation on various grounds. The oldest (Hume’s and Bentham’s) is that animals can suffer, and that it is cruel to inflict suffering. (Descartes had cleared the way to animal experimentation by holding that animals are soulless machines.) Another is that, by being compassionate with animals, we may become more compassionate with people. A third is that there are no lower animals: that all are equivalent and with the same rights as us. The first two reasons are valid but not compelling: all they tell us is to refrain from unnecessarily cruel experiments — e.g. that whenever possible the experimental animal should be anesthesized. The third reason is invalid because (*a*) the animals endowed with more highly evolved central nervous systems are objectively superior to the others in matters of cognition and emotion — in particular their sufferings can be the more intense, and (*b*) rights are not innate or natural but either conquered or given by humans. (See Landsell 1988.)

If the animal rights activists had their way, research and teaching in zoology, medicine, veterinary medicine and other fields would grind to a halt. In particular, the laboratories of experimental ethology, where the optimal conditions for domestic animal welfare are studied, would be shut. Consequently the well-being of all animals, in particular of humans and farm animals, would be considerably lowered.

The morally right thing to do with regard to animal experimentation is (*a*) to stop mindless experimenting, of which there is a lot, and (*b*) to review carefully the experimental designs, as well as the living conditions of the experimental animals, with a view to minimizing their suffering. As a matter of fact serious efforts are being made to improve these conditions. But the major source of useless experimentation, animal or otherwise, namely the dearth of clear original hypotheses, is still strong. This offers philosophers of science an opportunity of being kind to animals by way of being ruthless to mindless empiricists.

Finally comes the turn of genetic engineering or molecular biotechnology. Humans have manipulated genes, albeit indirectly, since the Neolithic Revolution that occurred between ten and fifteen thousand years ago. They have done so by artificially selecting animals, plants, and even bacteria. They have seldom if ever had any moral qualms about it, not even when they produced animals suffering from severe genetic handicaps. But contemporary genetic engineering poses new tough moral problems, because it can redirect bioevolution almost overnight and therefore without enough time to weigh unforeseen consequences. In particular, the bioengineer may unwittingly produce viable monsters, particularly micro-organisms, capable of wiping out entire biopopulations — e.g. our own.

Appalled by this possibility, a number of eminent molecular biologists and genetic engineers signed, and initially observed, the Asilomar Convention (1975), which involved a moratorium on genetic engineering until such time as a better grasp of the mechanisms of genetic change could be had. This proposal proved impracticable because genetic manipulation is a prime source of important fresh knowledge — and because its practitioners are curious, ambitious and competitive. ("If I quit the race others will win it".) The upshot is that most of them have decided to ignore the moral issues.

There is little that the moral philosopher can do in this case except (*a*) to ask the genetic engineers to use their knowledge only to design useful organisms (such as frost-resistant plants, and bacteria capable of synthesizing medical drugs) and to help repair genetic defects (such as schizophrenia), while refusing to use it in weapons design, and (*b*) to advocate strict patent regulations concerning new biospecies, to prevent health and environmental hazards.

2.3 *Nomoethics*

Let us call *nomoethics*, or its hybrid and cacophonic synonym *jusethics*, the intersection of moral philosophy and legal philosophy, i.e. the discipline that studies the moral problems encountered in legal theory and practice. The very first problem of nomoethics is that of its very existence and, more generally, that of the relation between ethics and legal philosophy, and the corresponding relation between morality and the law. There are several possible views on this matter: See Figure 8.2.

According to legal positivism (Kelsen 1949) morality and the law are mutually disjoint, whence their corresponding theories are disjoint as

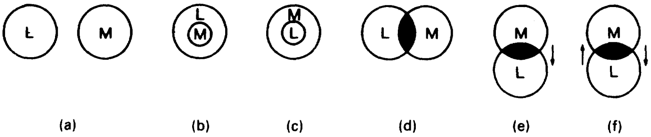


Fig. 8.2. The possible relations between the law (L) and morals (M), hence between legal theory and philosophy, on the one hand, and ethics on the other. (*a*) Legal positivism: L and M are disjoint. (*b*) contractualism: M is included in L. (*c*) Natural law: the right legal system is contained in M. (*d*) Common sense view: L and M have a nonempty intersection. (*e*) The high-minded view: the law should be dominated by morality. (*f*) The realistic view: L and M interact. Sometimes morals inspire legal reforms, at other times the latter alter the prevailing morality.

well. They could never come into conflict with one another, and there is no point in subjecting a legal code to moral criticism. Kelsen's motive for advocating this divorce between the law and morality was to turn jurisprudence into a rigorous discipline — a branch of logic free from metaphysics, psychology, politology and history. The resulting "pure" theory of law consecrates any consistent legal code, no matter how monstrous it may be, and it ignores the beneficial influence that morals have exerted on the law over the centuries.

Legal positivism is wrong-headed because the whole point of jurisprudence, as of ethics, is not argument for argument's sake but the content or meaning of value judgments and moral norms: logic is only a tool for analyzing them (Bruera 1945). By defining rights and duties in terms of legal codes, logical positivism embraces radical ethical relativism. It does not allow us to ask whether a given legal code fails to enshrine certain moral rights or duties, and whether we have any (moral) right to advocate law reforms on moral grounds. To the legal positivist the expression 'unjust law' is a contradiction in terms and therefore senseless. To anyone else, in particular a legal reformer, it makes perfect sense: A law is (morally) just if and only if it matches the corresponding moral right or duty.

A second view on the relation between the law and morals is that the former includes the latter. This is, presumably, the position of a radical consistent contractualist, namely that ethics is a branch of contract law. (Recall Ch. 7, Sect. 2.4.) This view is just as unhistorical and morally corrosive as legal positivism. Actually it is even worse, for at least Kelsen allowed for psychological conflicts between adherence to the law and observance of moral principles. The contractualist despises any moral norms that do not derive from contracts. All he wants is for the parties to sign on the dotted lines.

The third view, that the law is or ought to be included in ethics, does not seem to have been defended, not even by the natural law school. Moreover it is indefensible, for a legal code is nothing but a set of permissions and prohibitions, whereas a moral code may support us in our pursuit of justice, and an ethical theory may justify such pursuit; moreover a moral code may goad us to perform supererogatory actions and to refrain from immoral practices that are not penalized by the law.

The fourth view is that morality and the law are on the same footing but have a large overlap (Hart 1961). The evidence for this view is abundant: (*a*) every legal code contains or presupposes some moral norms (sometimes called 'frame' or 'constitutional' principles); (*b*) every legal code contains or presupposes a number of general principles that are moral as well as legal, such as "Innocent until proved guilty", "No crime without law, and no punishment without crime", and "No retroactivity"; and (*c*) there are legal wrongs that are not moral wrongs (e.g. the breach of a rental lease due to lack of funds), as well as moral wrongs (e.g. failures of reciprocity) that may not be subjected to legal sanction.

The fifth view is an elaboration of the former. It holds that morality and the law (hence ethics and legal philosophy as well) intersect in such a manner that the former has or ought to have the upper hand. In other words, the law should conform to morality, not the other way round. The virtue of this view is that it encourages law reformers to match (right) moral norms. For example, warmongering, the commodification of certain goods (e.g. blood), and the broadcasting of information concerning a reformed criminal's past, ought to be criminalized. However, the view has a flaw: it does not cover the case when a progressive legislature passes a law contravening an obsolete moral precept. For example, some parliaments have abolished the death penalty despite the opposition of public opinion, which clung to the *lex talionis*.

The sixth and last view is that there should be a vigorous interaction between morality and the law, hence between ethics and legal philosophy as well. We espouse it for scientific and practical reasons. The former because it matches what we know about the coevolution of morals and the law. (Recall the outlawing of slavery, serfdom, and torture, as well as the large body of law that accompanies the relief state, and the emerging environmental law.) The practical advantage of the interaction view is that it turns law makers into moral reformers, and it gives moral reformers a say in law reform. Such interactions should contribute to moralizing the law and institutionalizing morals while keeping both evolving along with society.

A good test case is the proposal to decriminalize the sale and use of hard drugs such as heroine and cocaine. The only reason for this proposal is a practical one, namely that such measure would reduce dramatically the price of such drugs and thus undermine organized crime and reduce the number of crimes committed by drug addicts. This is uncertain: there could still be drug cartels, as there are oligopolies and even monopolies in most sectors of a capitalist economy. The only practical solution is to have the state hold the monopoly on hard drugs, and put their distribution in the hands of state hospitals, which can control drug use and persuade some addicts to undergo therapy — as is the case in the United Kingdom and other countries. Besides, there is a powerful moral reason for not allowing any "hard" drugs to be dumped on the free market, namely this. If a vice is not punished it is no longer seen as a vice. Thus smoking was tolerated until a number of municipal governments outlawed it in public places, thus making it clear that smoking in public is antisocial, hence immoral. In short, morality and the law should go hand in hand.

So much for the relation between ethics and legal philosophy. Let us now examine a few specific and topical problems, starting with the question 'What kind of entity can stand trial in a court of law?'. This is a nomoethical question because it involves the notion of personhood. In the Middle Ages not only people but also animals, such as rats and ants, could be prosecuted; nowadays private companies, government departments and other social systems are regarded as legal persons. And some radical environmentalists hold that natural systems, such as forests and rivers, ought to be treated like persons and defended in court by "guardians" against polluters, loggers, and others (Stone 1974).

Only individual human beings can stand trial because they alone are moral agents, hence capable of infringing rights or duties. Companies and governmental services cannot be brought to trial: only their representatives can, and they should bear the responsibility for their own actions. This point is important: guilt and merit, like vice and virtue, are individual, not collective. Hence punishment and reward should never be collective. Thus, the German people was not guilty of Nazism, even though most Germans seem to have been enthusiastic Nazis as long as they were victorious. As for natural systems, such as rivers and forests, they have neither rights nor duties, so they cannot stand in a court of law. But everyone of *us* has the right to make correct use of them and the duty to protect them from ecovandals.

Most nomoethicians have focused on criminal law, which raises a number of tough philosophical problems centered around that of individual liberty. (See Feinberg 1984—1986.) This specialization has been double-edged. On the one hand it has led a number of philosophically minded lawyers and law-makers to advocate the repeal of harsh punishments and even the replacement of the very idea of retributive or punitive justice with that of corrective (or reeducating and rehabilitating) as well as reparative (or compensating) justice. But on the other hand it has, at least in the hands of the libertarians from Mill (1859) on, stressed the importance of rights to the detriment of duties.

The very first nomoethical problem raised by criminal law is that of elucidating the concept of moral crime as different from that of legal crime, or crime punishable by law. We stipulate that a *moral crime* is either (*a*) an action that interferes with someone else's exercise of her moral rights or duties, or (*b*) the failure to do one's moral duty. This definition covers crimes of omission as well as crimes of commission. Ontologically, only the latter are events or processes. Yet some philosophers have proposed unifying the two by stating that, if X refrains from doing Y, X makes the *negative* event not-Y happen. But this is verbal trickery, for the notion of a negative event is self-contradictory: there is no event if nothing happens. (Recall Vol. 3, Ch. 5, Sect. 1.) The person who neglects his duties is not acting but he allows certain undesirable events to happen: he is to blame for the events which he could have prevented.

Whether moral or legal, of commission or of omission, a moral crime may be environmental, biological, economic, cultural, or political. Until recently custom and legislation ignored environmental crimes, such as the dumping of toxic wastes. And even now most cultural crimes, such as playing a ghetto-blaster and publishing lies, go unpunished. In a just society most moral crimes are criminalized, and all legal crimes are morally condemned. In our societies a number of moral criminals are awarded decorations or positions of power.

Most criminal codes presuppose the general or frame principle included in the *Déclaration des droits de l'homme et du citoyen* of 1789: "The law ought to prohibit only actions harmful to society". Accordingly self-destructive behavior would not be a sufficient warrant for state interference. Yet all of the states in civilized nations practice what has been called 'legal paternalism', abhorred by the libertarians. In fact they prohibit voluntary enslavement and contracts without a dissolution clause, as well as drug abuse and self-mutilation; they force us to be inoculated against infectious diseases, to wear seat belts, and to send our children to school; in many cases they also force us to make contributions to our pension plans, to refrain from smoking in public places, and so on and so forth.

As a rule only the wealthy object to legal paternalism and insist that liberty is more valuable than welfare. Most of us have the converse preference and some of us advocate a welfare state with freedom and without alms. (See Ch. 11, Sect. 3.) Almost everybody understands that, when harming ourselves, we may inconvenience or even hurt others. For example, the reckless car driver puts the lives of others at risk, he taxes the public road security and the health systems, and increases the automobile insurance premia. However, any state protection measure should be introduced democratically and it should not hinder private initiative. Wherever the state becomes responsible for everything, the individual abdicates responsibility and thereby ceases to be a moral agent, as a consequence of which the law displaces morality.

Radical libertarianism is not the solution to the deficiencies and excesses of the relief state. For one thing it overlooks the immorality of leaving the weak unprotected on the excuse that legal paternalism infringes on individual liberty. Dworkin (1979 p. 96) put it succinctly: "better ten men ruin themselves than one man be unjustly deprived of liberty". But there is a third option: eleven reasonably happy and free human beings.

Secondly, radical libertarianism faces what may be called the Law and Order Dilemma. If the individual takes the law in his own hands, he becomes a vigilante, a lyncher, an outlaw: he ends up by destroying the very legal order that he claims to protect. If on the other hand he relies on an all-powerful law-enforcement body free from democratic control, he favors the emergence of a police state and a praetorian guard ever ready to use its power to squash all liberties. The solution to this dilemma is not to choose between liberty and welfare but to combine them. Only such combination can avoid the scourges of poverty, oppression, and corruption — which, alas, often come together.

2.4 *Business Ethics*

There are two traditional schools of thought concerning the relation between business and morals. One of them holds that business is amoral, the other that it is immoral. The former claims that economic relations are characteristically amoral, yet naturally constrained by competition. The (ideal) free market would be "the morally free zone", and "Were the world such a market, morals would be unnecessary" (Gauthier 1986 p. 13 and Ch. IV).

Anyone coversant with the business world knows that the amorality thesis is false. Firstly, every business transaction, even between criminals, calls for a minimum of honesty and trust. Even Milton Friedman, the well-know advocate of big business, when stating that the sole social responsibility of business is to make profit, added the caveat: "So long as it [. . .] engages in open and free competition without deception or fraud" (Friedman 1962 p. 133). Morality precedes agreement not the other way round (Ch. 7, Sect. 2.4). When this is not the case the firm and even the entire economy may be at risk: in the long run dishonesty is bad business. The banker Felix Rohatyn (1987) predicted the stock market crash of October 1987 as due in part to the sharp practices of a number of corporation managers and stock brokers. (More on the market and morals in Knight 1935, A. Buchanan 1985, Koslowski 1986.)

Secondly, high technology is not enough to guarantee high product quality: honest workmanship and pride in a job well done are necessary as well. Thus the phenomenal success of the Japanese economy is not just due to its imitation of Western technology, for other nations too have practised technology transfer and failed. The success is due to a combination of technology with economic policy (strong state interference) and morality. The moral factor is this: The social behavior of the typical Japanese is ruled by self-restraint, respect for others, and the sense of duty, rather than by self-interest. (See e.g. Sen 1987.) Japanese capitalism flourishes largely for having tapped a traditional morality rooted in conditions of scarcity and crowding.

So much for the credit side of capitalist business ethics. On the debit side the following points have been noted by uncounted observers and analysts: (*a*) the market is a rat race and as such is immoral, for in it the weak succumb to the strong, who are not necessarily the better; (*b*) unless checked by labor unions and government departments, the entrepreneur may have no qualms in exploiting his employees, cheating his customers, or contaminating the environment; (*c*) a substantial fraction of the industrial output consists of morally objectionally items, such as weapons, sports cars, electric guitars, drugs, junk food, war toys, and pulp magazines; (*d*) advertisement is designed to increase consumption and much of it is deceptive; (*e*) economic power corrupts just as much as political power: witness white collar crime and the bribing of public servants by companies bidding for government contracts; (*f*) international competition has often led to trade wars and even armed conflicts: witness the colonial wars, notably the First World War. In short, the economy is not without morals — but it is not a morality school either. (See a fair sample of views on business ethics in Snoeyenbos, Almeder & Humber Eds. 1983.)

The above mentioned flaws of the business world have been noted by the second traditional school of business ethics, namely the one holding that business is inherently immoral, so that the expression 'business ethics' is a contradiction in terms. Paradoxically this view is held not only by most Marxists but also by the Christian and Islamic advocates of a return to the Middle Ages, when prices and interests were strictly regulated — whereas the exploitation of serfs and the sacking of "enemies" were not.

Neither of the two extreme views, that business is amoral, and that it is immoral, is adequate. The truth is that (*a*) there is no social — in particular economic — behavior without some moral rules (Ch. 4, Sect. 3.3); (*b*) one cannot pursue one's own interests without any concern for other people's interests, for no one is fully self-reliant; (*c*) every modern economy, capitalist or otherwise, is constrained by laws and regulations aiming at protecting trade, credit, public health, and the environment, by penalizing moral crimes such as fraud and the breach of promise; and (*d*) any firm that wishes to stay in business must observe a minimal moral code, for otherwise it will not earn the trust of its business partners and customers.

As a matter of fact this third view, that business has or must have a moral code, is quickly gaining ground, as may be seen from the *Harvard Business Review* and the *Journal of Business Ethics.* A new code of business ethics is being fashioned and adopted by an increasing number of highly successful American companies (*The Economist*, July 2, 1988, p. 58). Its contrast with the traditional one is summarized in Table 8.3.

In short, whereas the traditional business morality was intended to advance only the interests of the shareholders of a firm, the newly emerging one is supposed to advance those of all its "stakeholders", i.e. all the people likely to be affected by the company's activities. This is certainly in keeping with the universality requirement of all the altruistic ethical theories (Ch. 7, Sect. 3). Still, a number of negative moral features remain: the unequal distribution of the surplus, the domination of the weak (individuals, firms, nations), the promotion of consumerism and waste, the unfettered exploitation of non-renewable resources, and so on. Sweeping reforms will be needed to correct these serious flaws without at the same time suffocating individual initiative and eliminating the creative aspect of competition. We shall address this problem in Ch. 11, Sect. 3.

2.5 *Political Ethics*

The views on the morals of politics are similar to those concerning business ethics. Indeed opinion is divided into amoralism, immoralism, and moralism. Among the amoralists we find not only Machiavelli and

TABLE 8.3. New *vs.* old business morals: a sample of norms.

Old guard companies	Vanguard companies
The only social responsibility of business is to make profit (Milton Friedman).	"The business of business is serving society, not just making money". (From the constitution of the Dayton Hudson corporation.)
Maximize the benefit/cost ratio.	Optimize the benefit/cost ratio.
Milk and, if need be, cheat the consumer.	Seek consumer satisfaction and loyalty.
Disregard environmental concerns.	Respect the environment.
Disregard worker satisfaction: seek only productivity and obedience.	Seek worker satisfaction and loyalty.
Confront and if possible destroy or corrupt the labor organizers.	Bargain and cooperate with the labor unions.
Practice authoritarian management.	Practice participatory management.
Oppose government intervention when prosperous, demand it when in the red.	Accept and if necessary seek government intervention to secure fair business practices.

Lenin but also most politologists, particularly those who swear by game theory. The second group is the large body of public opinion disenchanted with the contrast between election campaign promises and the realities of power. The third group is made up of reform politicians and theorists. The most radical among them is still Aristotle, who held that ethics is a part of what we now call political science (*Nichom. Ethics* I, 3, 1094b).

Every one of the three views contains a grain of truth. Amoralism is correct with regard to the purely technical ("mechanical") aspects of the organization of political activities and public services, which can be entrusted to experts. But amoralism is incorrect in everything else. For example, a party program and a government policy are bound to have moral components because their implementation is likely to favor some social groups at the expense of others: just think of the different ways tax revenues can be used — e.g. to buy health and education or to buy security for the powerful. In short, political amoralism is just as false as business amoralism. It is also morally wrong for holding that the end justifies the means.

The immoralists, i.e. those who believe that politics is inherently and therefore incurably immoral, make their case by listing the varieties of corruption in politics: buying votes, selling government jobs, taking bribes, lying, oppressing, waging offensive wars, invading foreign countries, using the armed forces to quell social discontent, and so on. But the immoralist is wrong in holding that all politicians are corrupt: suffice it to mention Pericles, Gandhi and Palme, not to mention the millions of devoted rank and file political activists who work selflessly for a cause they believe to be good. The immoralist is also wrong in holding that politics cannot be cleansed. In fact democratic politics is a morality school, for it teaches cooperative habits, rational and respectful dialogue, and ideals of public good. Therefore immoralism is false. Worse: it is morally and politically wrong, because it erodes the trust in democracy and paves the way for tyranny.

Let us now deal quickly with a few political questions involving morals. Our first question is: Why bother to participate in politics? The answer is simple: Where people do not participate in politics, this becomes the monopoly of a political class that may not defend the public interest but some special interest, e.g. its own. If we value genuine democracy we must bother to vote, attend meetings, take part in political campaigns, and be ready to serve, however sporadically, in public offices.

Voting, the citizen's minimal obligation in a political democracy, has been the subject of ingenious theoretical studies over the past two centuries. Politologists, from Condorcet to Arrow and beyond, have concluded that no voting decision procedure is good. These "impossibility" theorems are suspect because they all presuppose that every voter attempts to maximize personal gains. In fact people vote for a number of reasons: not only to defend their own interests (as they perceive them), but also out of a sense of civic pride, out of devotion to an ideology, because it feels good to exert a right, and even better to belong to a group, or merely to comply with the law. People vote for a variety of reasons and sometimes at the cost of discomfort or even the risk of physical violence. And, even though no voting procedure is perfect, some procedures (e.g. the French one) are better than others, and any voting is better than none.

Our second question is: Why tell the truth in politics? Why indeed, when lying is so much more profitable in the short run? The moral motivation is obvious. But there are prudential motives too. One of them is that only morons or cynics vote a second time for a convicted liar, or continue to support a government that is caught lying once and again. Another rationale is that democracy involves cooperation in some respect, which requires trust, which calls for truthfulness. In conclusion, in the long run political forthrightness is always the more convenient policy as well as the only moral one.

A related problem is that of the moral status of political compromise. Politics has been called the art of compromise; it is also the art of detour and of waiting for opportunities and making the most of them. Without practising these arts there can be no hope of attaining power or of retaining it. But are such stratagems moral? This depends on whether or not any basic *moral* principle is being sacrificed in the compromise, detour, or postponement. Never mind the ideological tenets: if they were morally wrong to begin with, there is nothing wrong with giving them up. But of course if a politician does change his ideology he has the moral duty to say so. In short, *pace* Weber (1921), the politician may refuse to "sign deals with diabolical forces".

Our next subject is tolerance. The conservative view is of course intolerant. It is also the most natural: tolerance is learned in social intercourse, and political tolerance is a recent acquisition, and moreover not a universal one. It is not yet generally understood that diversity is good, provided it does not involve injustice, because it presupposes some freedom, it is a condition of progress, and it makes life interesting. Nor is it universally understood that uniformity requires submission and results in stagnation. Even in democratic societies there is strong pressure to conform, and original people are often seen as trouble makers.

Even people who tolerate other people's customs and ideas are often politically intolerant, i.e. conservative — of the right or of the left. Hence they are ready to accept censorship and repression. But there are also liberals ready to protect the right to hold and advocate any doctrine, however intolerant — e.g. racism, aggressive nationalism, and economic elitism. Witness the American Civil Liberties Union. To be sure such open-mindedness and generosity are admirable — but also suicidal. It calls to mind the hospitality extended to the convicted arsonist or criminally insane.

If a good society is to be built or preserved, we must protect the creative nonconformist while controlling and reeducating those who engage in antisocial behavior, be it brawling or rioting, stealing or murdering, warmongering or gunrunning. We should adopt what may be called the principle of *moderate tolerance*: "Tolerate anything except what is sure to hinder the satisfaction of the basic needs and legitimate aspirations of everyone". Once again, the standard is the supreme maxim "Enjoy life and help live".

Patriotism is usually regarded as a civic virtue. Is it really? The answer depends on the kind of patriotism concerned: extreme, middling, or moderate. Extreme patriotism (jingoism, chauvinism) is the same as aggressive nationalism. It boils down to the simple-minded slogans "My country right or wrong", and "My country before any other". The first slogan is immoral for erasing the difference between right and wrong; and it is silly for blinding us to our country's shortcomings, thus blocking any attempt to correct them. And the second slogan is immoral for inciting to war and conquest. In short, extreme patriotism is not virtuous.

Middling patriotism can be summed up in the formula "Everyone ought to advance the interests of his own country". It does not involve blind conformism and it does not incite to international aggression. But, since it does not mention the interests of other countries, it tacitly condones international injustice. Finally, moderate patriotism is encapsulated in the formula "Everyone ought to advance the interests of his own country as long as these do not conflict with the rights of others, starting with the rights to survival and self-determination". Evidently only this kind of patriotism is consistent with our moral doctrine.

What about violence? Nothing works like fear, particularly the fear of violence, to tame people and force them to abdicate their moral and civic responsibilities. Political violence is morally wrong for violating basic rights, such as the right to life; and it is prudentially wrong because it tends to be self-perpetuating. A good society is internally and externally peaceful: its citizens, particularly its political leaders, use only peaceful means to resolve conflicts. (Such means may go from talks and bargains to mass rallies and passive resistance.) In such a society violence is only justified to repeal individual or collective aggression.

The only effective and lasting way of coping with political violence is by preventing it. This goal is not achieved by strengthening the so-called security forces, because they have a vested interest in keeping conflicts alive, and they may end up by controlling the government. Political violence can only be prevented by (*a*) minimizing the sources of civil and international conflict, (*b*) promoting interpersonal and international cooperation, and (*c*) banning all weapons, from hand guns to nuclear bombs, the world over.

The most serious threats to internal and international peace are posed by the empire builders, the arms industry, the armed forces trained and armed for aggression, and the jingoistic and militaristic ideologists. Some of the latter have had the cheek to attempt to justify war on moral grounds. Thus the first fascist claimed that war brings out the best in man, whence "fascism believes neither in the possibility nor in the utility of perpetual peace" (Mussolini 1932 p. 47). Bergson (1932 p. 26), writing at the same time, knew better: In war "Murder and pillaging, like perfidy, fraud and lying, do not just become licit but also meritorious. Like *Macbeth*'s witches, the belligerents will say: Fair is foul, and foul is fair". Even as he drew his country into World War I, President Wilson admitted that, when at war, people lose their heads and "stop weighing right and wrong". In short, war "makes clear and honest thinking impossible" (Mencken 1956 p. 227). So does military training, for it involves self-abasement, the repression of moral feelings, and abdication of moral responsibility.

Finally, are there just wars? It is usually taken for granted that all national independence wars, as well as all the wars waged in self-defense, are just. (See e.g. Walzer 1977, Calvocoressi 1988.) But no war *can* be just, because it is mass murder and destruction; at most, one of the *sides* in a war, namely the victim of aggression, can be morally right — provided it limits its actions to repealing aggression and, in case of victory, does not take advantage of the defeated. It is never morally right to initiate an armed conflict, not even in defense of a good cause, because (*a*) the initial goal is inevitably replaced by that of winning at all costs, and (*b*) the cost in lives and resources, as well as the ultimate outcome of the conflict, are hardly predictable. In short, there are no just wars, not even those allegedly waged "to end all wars".

2.6 *Summary*

The exercise of any profession is bound to affect the welfare of some people, hence it must be encouraged or restricted by moral norms. As a matter of fact every profession is ruled by a more or less explicit moral code of its own. A moral code is the more complex, the stronger the social impact of the professional activity.

The only professional ethics that has been explored in some detail is bioethics. However, excessive attention has been devoted to comparatively minor issues such as abortion and euthanasia, at the cost of the neglect of the bulkier issues of public health and preventive medicine. In view of the power attained by surgery, drug treatment and behavior therapy, the Hippocratic Oath ought to be revised. We need a code based on the value judgment that not just life, but life worth enjoying, is the highest value, hence the one to be protected by all means.

Nomoethics, or the discipline that studies the moral problems encountered in legal theory and practice, admits that, far from being mutually disjoint, morals and the law overlap partially. Moreover, they interact, and they should do so more vigorously. One of the central problems of nomoethics is that of distinguishing moral from legal crimes. Classical liberalism holds that the law ought to prohibit only actions harmful to society. By exerting legal paternalism, the relief state goes far beyond the liberal precept.

There are three views concerning business ethics: amoralism, immoralism, and moralism. The first two are not only morally objectionable but also at variance with business practice. Indeed, every business transaction calls for a minimum of reciprocity and trust. The old business ethics, or rather nonethics, centered on the maximization of benefits precept, is on the way out. New winds are blowing in business theory and practice. It is being increasingly understood that management has to serve the interests of all the "stakeholders" of the company rather than just those of its shareholders. Still, a number of serious moral issues remain unresolved.

In political ethics too there are three schools: amoralism, immoralism, and moralism. The former is wrong if only because every state policy is likely to affect, favorably or unfavorably, the welfare of some people. Immoralism does not fit the facts either: There have been uncounted honest political leaders and civil servants. Any person active in politics faces a number of moral problems, which moral codes and ethical theories ought to help solve. The most serious of all such problems is the legitimacy of violence. It is proposed that violence is only justified in repealing aggression, and that war is the greatest of all crimes. *Si vis pacem para pacem.*

CHAPTER 9

METAETHICS

In Antiquity ethics was regarded as the study of virtue and the good life. From the beginning of the modern period the task of moral philosophers became that of proposing and discussing theories about the moral (as distinct from the legal) rights and duties of all human beings — or at least of all white and adult males. Whether ancient or modern, traditional moral philosophy was directly relevant to real life — so much so that it was also called 'practical philosophy'. But admittedly it was conceptually rather sloppy. The so-called linguistic turn (roughly 1930—1960) did something to correct this shortcoming by focusing on the analysis of key moral concepts, such as those of person, action, responsibility, and norm. By and large, the study of moral problems and ethical doctrines gave way to the study of the "grammar" of moral terms, i.e. the way we ought to talk about morals if we wished to make sense — at any rate to fellow academics in the Anglo-Saxon world. Ethics was displaced by metaethics, and action by talk.

The so-called linguistic turn of moral philosophy proved to be ambivalent. On the one hand it had a healthy effect in that it swept away much obscurity and confusion. Moral philosophy became transparent — just as transparent as the void. Philosophers became very careful about words — so cautious that they said very little, if anything at all, about real-life moral issues. The negative effect of the linguistic turn was precisely this, that it gutted moral philosphy. The model moral philosopher was the gutless academic who avoided moral problems like the plague and remained always *au dessus de la mêlée.* He posed as a technician servicing a machine (language) rather than its user (the moral agent). He lived in a comfortable verbal cocoon that protected him from moral conflicts: his job was to analyze, not to moralize. Worse, his technique proved to be not that sophisticated after all. In fact it was either ordinary language analysis *à la* Wittgenstein II or some logic of norms that raised more problems than it solved.

Since the mid-1960s the pendulum has started to swing in the opposite direction. A new generation of moral philosophers, concerned about real-life moral issues, has given new life to ethics without, however, returning to the obscurities and confusions of preanalytic philosophy. They have started to wrestle with such moral conflicts as private interest-public interest, peace-national security, industrial growth-environmental protection, liberty-big government, and technological development-welfare. The typical contemporary moral philosopher asks not only what the words ‘right’ and ‘duty’ *mean*, but also what right and duty *are*. He is interested in facts as well as in words; in ethics and its subject matter (morality) as well as in mataethics; in political philosophy as well as in moral philosophy. This reorientation of moral philosophy is all to the good. Still, it is only a beginning.

In the present chapter we shall examine the nature of ethics and its relation to the real world: we shall engage in metaethical analysis as a complement of, but not as a substitute for, ethics.

1. SCIENTIFIC AND PHILOSOPHIC ETHICS

1.1 *Ethics as Science and as Philosophy*

In ordinary language and in linguistic philosophy the words ‘ethics’ and ‘morals’ are often used as synonymous. In contrast we have been distinguishing morality, as a psychological and social fact, from its study; in particular, we have distinguished moral codes from ethical doctrines. (See also C. O. Bunge 1919.) Actually we must introduce an even finer distinction, namely between (*a*) *moral facts*, such as someone helping a passer-by in distress, (*b*) *moral codes* or systems of moral norms, (*c*) *ethical doctrines* or theories about such codes, and (*d*) *metaethics*, or the study of ethical concepts, principles, and theories. See Figure 9.1.

The *semantic* relation between each of the higher levels distinguished above and the next lower level is that of reference, studied in Vol. 1, Ch. 2. Metaethics refers to ethics, the latter to morals, and these to moral facts or moral agents. Hence ultimately metaethics refers to moral facts or agents. (The composition of two or more reference functions is itself a reference function. Recall our analysis of metanomological statements in Vol. 7, Part 1, Ch. 2, Sect. 1.2.) In other words, the referents or units of moral discourse in either of the three higher levels are ultimately persons or groups. (The higher animals count as persons because they have higher mental capacities: recall Vol. 4, Ch. 4, Sect. 5.2.)

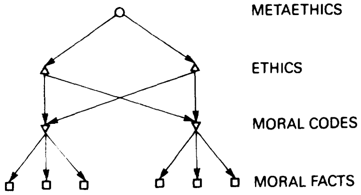


Fig. 9.1. Four levels of moral and ethical items.

The *epistemological* relation between any of the higher levels and the next lower one is that of studying or investigating. In particular metaethics analyzes, systematizes and evaluates ethical ideas; and ethics studies, systematizes and evaluates morals. But morals regulate moral facts (or rather agents) instead of studying them. The difference between metaethics and ethics is analogous to the difference between the philosophy of science and science. And the difference between ethics and morals is similar to the one between factual science and reality. Morals are "out there" in social behavior as well as "in here", in the individual's conscience (or rather in his brain).

Metaethics and ethics can be descriptive, prescriptive, or both. That is, they may study their object or regulate it. In particular, metaethics may rule ethical theories, which in turn may regulate moral codes: metaethics is a sort of conscience of ethics.

Ethics may be scientific, philosophical, or both. On the one hand, *scientific* ethics, or moral science, studies moral behavior from a psychological, sociological or historical viewpoint, whereas *philosophical* ethics studies the same subjects from a logical, epistemological, methodological, ontological or axiological viewpoint. On the other hand, metaethics is a strictly philosophical discipline even though it ought to admit, nay seek, scientific inputs. The union of philosophical ethics and metaethics may be called *moral philosophy*.

The very first task of metaethics is to characterize ethics. We shall do so following the general definition of research field proposed in Vol. 6, Ch. 14, Sect. 1.2. We stipulate that *ethics* is representable by the ten-tuple

$$\mathscr{E} = \langle C, S, D, G, F, B, P, K, A, M \rangle,$$

where

C = the community of ethical investigators;

S = the society hosting C;
D = the referents of $\mathscr{E}$;
G = the general outlook or philosophy of $\mathscr{E}$;
F = the formal background or collection of formal tools employed in $\mathscr{E}$;
B = the specific background of (information relevant to) $\mathscr{E}$;
P = the problematics or collection of problems of $\mathscr{E}$;
K = the fund of knowledge or ethical tradition;
A = the aim of $\mathscr{E}$;
M = the methodics or collection of methods of $\mathscr{E}$.

Our distinction between scientific and philosophical ethics, sketched a moment ago, may now be spelled out in some detail:

	SCIENTIFIC ETHICS	PHILOSOPHICAL ETHICS
C	The *scientific ethics community* is composed of the psychologists and social scientists who study current or past morals in a scientific fashion;	The *philosophical ethics community* is composed of the philosophers who study morals in a philosophical manner;
S	the *society* hosting C is capable of encouraging or at least tolerating the professional activities (research, teaching, publishing, participating in meetings, etc.) of the members of C;	
D	the *domain* or *universe of discourse* of $\mathscr{E}$ is	
	the collection of persons and social groups as regards their moral beliefs and codes of social behavior;	the collection of moral codes and the family of ethical doctrines — past, present, or future;
G	the *general outlook* or *philosophical background* of $\mathscr{E}$ is	
	the one common to all factual sciences: a naturalistic ontology, a realistic epistemology, and the ethos of the search for truth;	a family of explicit or tacit semantical, epistemological, ontological and axiological views, as well as the ethos of the search for truth;
F	the *formal background* of $\mathscr{E}$ is the collection of logical or mathematical concepts, theories, and methods taken for granted and utilizable in the course of ethical inquiry;	
B	the specific background of $\mathscr{E}$ is a collection of data, hypotheses, theories, and techniques, borrowed from psychology, the social sciences, and philosophy, possibly relevant to the conduct of ethical inquiry;	

P	the *problematics* of $\mathscr{E}$ consists of	
	problems concerning the description, explanation and evaluation of moral codes;	problems concerning the logical form, meaning, consistency, and validity of moral codes, as well as the analyses and systematization of moral ideas;
K	the *fund of knowledge* of $\mathscr{E}$ is the collection of	
	propositions (in particular data and hypotheses) and procedures (in particular techniques) considered correct and obtained by members of *C* at previous times;	propositions, still regarded as worth discussing (even if thought to be incorrect), and obtained by members of *C* at previous times;
A	the *aims* or *goals* of the members of *C* are	
	to enrich our knowledge of morals, to understand (in psychological or sociological terms), evaluate, and help perfect them;	to enrich our knowledge of moral concepts and norms, to understand (with the help of conceptual analysis), evaluate, and help perfect them;
M	the *methodics* of $\mathscr{E}$ consists in	
	conceptually and empirically scrutable (analyzable, criticizable, checkable) procedures, in the first place the scientific method.	conceptual (in particular logical, mathematical and semantical) analysis and rational reconstruction (theory building), aided by imagination and scientific findings.

By definition, scientific ethics is included in the union of psychology and social science, whereas philosophical ethics is part of philosophy. However, the former makes use of concepts elucidated by the latter, which in turn may utilize findings of the former. Consequently scientific and philosophical ethics, far from being mutually disjoint, have something in common. If the former is a genuine science, every one of its components will change, however slowly at times, as a result of scientific research in its own field as well as in the relevant sciences and in philosophy. If approached scientifically, philosophical ethics will interact vigorously with scientific ethics, so that it will change over time, however slowly, in response to such interactions as well as to internal research. Given the close relationship between scientific and philosophical ethics, we define *ethics* as the union of its scientific and philosophical aspects. (Strictly speaking we must take the pairwise unions of all ten components of the two ten-tuples.)

Ethics is not limited to the study of moral norms: it must also systematize or codify them, evaluate them, and check whether in fact their observance contributes to the explicit or tacit goal of morality, e.g. the satisfaction of basic needs and legitimate wants. Furthermore, ethics is not only descriptive but also prescriptive: it attempts to build or reform viable systems of moral norms. In our view any such system should be (*a*) internally *consistent* (non-contradictory), (*b*) *compatible* with the relevant natural and social sciences, (*c*) *viable* — i.e. it must be possible for an ordinary human being to live up to it, and for a society to adopt it; (*d*) *balanced* — i.e. it must combine rights with duties; (*e*) *flexible*, i.e. it must make room for extenuating and aggravating circumstances; and (*f*) *humane*, i.e. designed to facilitate the pursuit of well-being. Any person or group capable of coming up with such a moral code deserves being called an "inventor of the future" (Augusto Bunge 1915).

1.2 *Scientific and Axiological Basis*

In the preceding chapter we argued that ethics has a number of scientific and philosophical inputs, among them social science and value theory: Recall Fig. 8.1. In the preceding section we characterized ethics through a list of ten items, one of which is the specific background B, i.e. the body of (scientific and philosophical) knowledge that ethicians take for granted or borrow. In particular, we claim that ethics presupposes, and partially overlaps with, both axiology and social science.

Our thesis is unpopular. The most popular views on the matter are that ethics is either part of theology (e.g. Augustine and Aquinas) or an autonomous discipline (e.g. Kant and Moore). Other views are that ethics is only based on value theory (N. Hartmann); that it is contained in value theory (R. B. Perry) or conversely (C. I. Lewis); that it is an offshoot of biology (E. O. Wilson), a part of psychology (M. Schlick), of sociology (E. Durkheim), or of political science (Aristotle, J. Bentham, A. Sen); finally, that ethics, together with the theories of decision and of games, forms the science of rational individual and collective choice (J. Harsanyi). Let us examine quickly these views.

The theological view of ethics is at variance with the whole of ancient Greek philosophy and with modern philosophy, starting with Hobbes and Spinoza. This is not to deny that some ethical theories are theological, and that the corresponding moral codes are religious. The point is that it is possible to deliberate on moral problems, and to reach correct decisions, without bowing to religious dogmas. Moreover, it may be argued that adherence to such dogmas implies abdication of moral responsibility and denial of accountability to others. (Recall Ch. 7, Sect. 1.1.)

This does not entail that ethics is autonomous. On the contrary, it depends on psychology, anthropology, sociology, and other sciences, for, after all, it studies social behavior, i.e. behavior capable of affecting conspecifics. Ethics also depends on value theory because all rational prescriptions are preceded by rational evaluations. A moral prescription that were not regarded as good for the purpose in question would be either arbitrary or perverse. The right is defined by the good not the other way round (Ch. 4, Sect. 1.1). See Figure 9.2.

The alternative views listed a moment ago are unacceptable for the following reasons. Firstly, ethics is based on more than value theory: it must make some use of scientific findings if it is to evaluate correctly the various moral codes, i.e. if it is to check whether they work in practice. A fortiori, ethics is not contained in value theory. Nor does the converse relation hold, for many a value judgment is not moral. As for the relations between ethics and science, it is not true that all of the former is contained in the latter: the inclusion relation holds only for the scientific part of ethics, not for its philosophical part.

Finally, ethics is not a component of a separate science of rational choice, because there is no such science. In particular, decision theory and game theory are not reliable guides for choice because (*a*) they

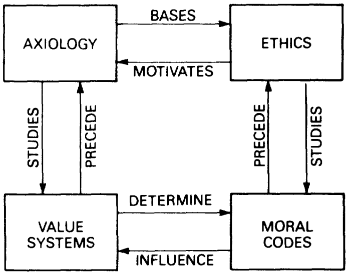


Fig. 9.2. Relations between ethics and value theory as well as between their referents.

contain no substantive knowledge and (*b*) they make use of subjective values and subjective probabilities (Ch. 3, Sect. 3.3). Rational choice, particularly when it is likely to affect others, is based on objective knowledge. For example, a decision to perform surgery must be based on the medical information that it is necessary to restore health, on the value judgment that health is preferable to illness, and on a risk assessment based on statistics. Another example: Think of the strategic analyst who employs game theory to find out whether a first strike with nuclear missiles would be convenient, and in doing so makes subjective estimates of the most probable number of casualties and uses a subjective estimate of the probability of retaliation. Some science, some morality!

1.3 *Constraints on Moral Norms*

Every explicit and viable moral code is constrained by (*a*) the nature of the world, in particular society, (*b*) a general frame or "moral constitution", and (*c*) a set of second-order rules (or meta-rules, or metanorms, or metaethical principles). The first or factual constraint is the collection of biological, psychological and social regularities and circumstances that determine whether a moral rule is necessary for some purpose or merely a pious hope.

The *general frame* or "moral constitution" (Kolnai 1978 pp 113 ff) is a set of general principles about man and the world: it is part of the general outlook *G* introduced in Sect. 1.1. A moral constitution is to a moral code what the constitution of a land is to its positive laws. Thus the frame of a theistic morality will contain the postulate that man is God's favorite creature, whence he must rely on divine providence and benevolence. On the other hand a naturalistic frame will include the assumption that man can only rely on himself and others, and is accountable only to himself and others. It is obvious then that the frame or constitution determines largely the content of a moral code.

The *meta-rules* of a normative system, be it moral, legal, or technical, are norms that regulate the content and scope of the substantive rules, i.e. the precepts that govern all of the actions of a certain type. The nomoethical principle *No crime without law* is a case in point. It is a principle about norms and their application. It does not tell us what the laws are nor what to do in a particular case — except to do nothing in case there is no (legal) crime. Two ethical examples are the Principle of Utility when applied to rules rather than acts, and Kant's requirement of universality of moral norms. Both constrain the work of the moral philosopher but neither tells us what to do in any particular situation. They are rules about rules not actions. See Figure 9.3.

In one of its versions, Kant's celebrated Categorical Imperative is the metanorm that all moral rules ought to be universal or universalizable. (This metaethical postulate has sometimes been mistaken for the Golden Rule, which refers to actions not rules. Kant's is clearly a "second-order principle": Broad 1930 p. 120.) Two reasons have been offered in support of it, one logical and the other moral. The former is that, if exceptions to a rule were admitted, the rule would not qualify as a principle or general law. However, this is not a valid reason, for one may state a universal law or rule in such a way that its antecedent covers the allowable exceptions. Example: "Unless knowledge of the truth is bound to harm the innocent, it ought to be told".

The second reason is that all human beings are equal and therefore deserve to be treated equally and, in particular, to be subjected to the same moral rules — which for Kant concerned only duties. This ground is shaky because, as a matter of fact, people are not equal, and one of the marks of a fair moral code is that it enjoins us to compensate for inequalities. For example, children, the aged, the handicapped and the sick deserve being treated better than able-bodied adults because they are at a disadvantage. It would be unrealistic and unfair to bestow exactly the same rights, and demand the same duties, regardless of needs and abilities.

This is not to deny the need for universal moral norms, such as "To each according to her needs, from each according to her abilities". Every viable moral code has some such universal norms. But every realistic moral code should also contain some *special* norms, applicable

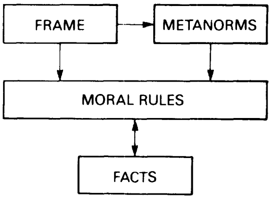


Fig. 9.3. Constraints on moral rules: factual, "constitutional", and metaethical.

to all human beings with similar needs and abilities and in analogous circumstances: one for normal children, another for retarded ones; one for normal adults, another for cripples; one for manual workers, another for intellectuals, and so one. These special norms should be superimposed on the general ones. Every such special rule would be *binding on everyone in the same category*. By proceeding in this way one obtains a *family of moral codes* that combine maximal universality with maximal flexibility. Universality satisfies the ideal of equality, and specificity matches the variety of persons and circumstances. See Figure 9.4.

1.4 *Norms and Normative Systems*

The very existence of norms and systems of such poses a number of interesting philosophical problems. One of them is ontological: What are they — conventions, laws, or a mixture of both? Another is logical: What is their logical form and how can we reason correctly with them? A third is of a semantical nature: What is their meaning and truth value, if any? A fourth one is epistemological: How do we get to know them and how are they related to descriptive knowledge? A fifth is methodological: How can they be validated or justified? In this and the following section we shall only take a peak at each of these questions.

Human beings satisfy laws of nature and invent, obey or break rules of social behavior. Whereas laws are invariant patterns of being and becoming, rules are historically changeable artifacts. (Vol. 5, Ch. 9, Sect. 3.2.) However, in morals and elsewhere not all rules are mere conventions. Some of them are *law-bound*, i.e. grounded on objective laws. Thus the rule of mutual help is law-bound insofar as (*a*) no individual is fully self-reliant, and (*b*) the persons who abide by the rule fare better than those who do not. We submit that all *valid* moral rules,

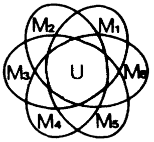


Fig. 9.4. The rosette of particular moral codes M_1 to M_6. Their intersection U is the universal moral code, that holds or ought to hold for every human being.

unlike most grammatical rules and many legal rules, are law-bound rather than conventional.

The distinction between valid and invalid moral rules is philosophical rather than scientific. Indeed, social scientists do not distinguish between morals and mores, between the right and the customary: their business is to ferret out the norms that regulate social behavior, be they moral, amoral, or immoral from a given point of view. Thus they are likely to treat the prevailing rules concerning truthfulness, loyalty, and the taking of life, on the same footing as the rules concerning kinship relations, property rights, or religious worship. On the other hand philosophical ethics, or moral philosophy, is supposed to propose grounds for or against moral rules: it draws a clear distinction between a valid moral rule and a mere custom. Thus a humanistic moral philosophy will condemn murder, whether individual or massive, even though in many countries murder is a growth industry. In sum, every socially accepted morality is a *mixture of genuine moral norms and customs*, some of them conventional. This explains why any given moral code takes on different forms in different social groups.

Moral and legal norms have traditionally been approached in an authoritarian way, i.e. they have been construed as permissions, prescriptions or proscriptions, no reasons or causes being offered. In particular, Kant thought that all moral norms are categorical (unconditional) imperatives of the form 'Do *x*!' or 'Don't do *x*!'. (A hypothetical or conditional imperative is of the form 'If *x*, do *y*!', or 'If *x*, refrain from doing *y*!'. 'Shut the door!' is a categorical imperative and 'If there is a draft shut the door!' a hypothetical one, though of course neither expresses a moral norm.)

It is mistaken to think of moral norms as orders, because many of them are injunctions, recommendations, or even proposals. Witness "Love thy neighbor" and the Golden Rule. A rational person will use imperatives only for issuing orders, not for couching moral norms: he will express the latter as *sentences*. Moreover, if he happens to be a consequentialist moral philosopher, he will prefer *conditional* sentences hinting at the reasons or causes for doing or refraining from doing whatever may be in question. For instance, instead of barking 'Don't litter!' he may prefer to say 'If you care for others, or wish to avoid being fined, do not litter!. And if he proposes this as a universal norm, he will state it thus: 'For every *x*, if *x* is a conscious person, and *x* cares for others or wishes to avoid being fined, *x* refrains from littering'. The antecedent of this general conditional is the *ground* or *reason* for its consequent. The advantage of the declarative over the imperative form is that sentences express more or less true or false propositions, hence they can be argued about with the help of ordinary logic, and tested with the help of empirical data. They cease to be dogmas to become testable hypotheses. (Recall Ch. 8, Sect. 1.1.)

An authoritarian ethics requires some logic other than ordinary logic in order to analyze moral principles and their relations, because imperatives do not satisfy the laws of deductive logic. This is why the standard logical analysis of norms is performed in terms of some system of *deontic logic*, which in turn is an interpretation or an extension of one of the 256 possible systems of modal logic. (See e.g. Lemmon 1965.) This interpretation construes 'p is possible' as 'p is permissible', or 'Pp' for short, where P is called a 'deontic operator'. A prohibition is analyzed as $\neg Pp$ (p is forbidden), and a duty as $\neg P \neg p$ (p is obligatory). The concept of a right does not occur in such a system: only gracious permissions and ungracious obligations and prohibitions occur in it. When the concept of a right is added (as in Kanger 1957), a far broader system results.

There are several problems with deontic logic. One is that, as we have just suggested, it takes it for granted that morality is a grim affair that has no room for enjoyment and aspiration. Second, it consecrates authoritarianism instead of inviting criticism and testing. Third, it does not involve a sound interpretation of modal logics. For example, whereas "p is necessary" implies p, its deontic interpretation "p is obligatory" does not imply p (or "it is the case that p"), for many an obligation is not met. Likewise, whereas p implies that p is possible, it is not true that p implies that p is legally or morally permissible, for one will commit actions that are not permissible. For this third reason Weinberger (1985) and a few others have concluded despondently that half a century of inquiry into the field of the logic of norms has ended in impasse.

But there is a fourth, perhaps even more radical, objection to the modal approach to norms, namely this. If a *proposition* were forbidden it would not occur in the moral or legal code in question: it might occur only in an *Index* of prohibited propositions. Only human *actions*, or those performed by animals or machines under human control, can be prohibited, obligatory, or permitted. In short, it makes sense to apply a deontic operator to human actions, not to propositions. And it makes even better sense to apply it to ordered triples of the form ⟨agent, action, code⟩. Example: “All and only able-bodied adults have the right to work according to the UN charter of human rights”.

We have moved from imperatives to propositions, and from permissions to rights; likewise we shall move from obligations to duties. The first move is purely formal, yet it has momentous logical and methodological consequences: it spares us deontic logic and it turns ethics from a set of dogmas into a system of testable propositions, the way Ingenieros (1917) wanted. The second move, from permissions to (moral) rights, and the third, from obligations to (moral) duties, weaken the traditional bonds between ethics on the one hand and theology and the law on the other. At the same time they reinforce the bond between moral norms and moral values, hence the difference between morals and mores. (A despot may have deprived me from exercising the right to do X, but if this right is moral I will be motivated to fight for it, and others will support my cause.)

The propositionalization of imperatives has been independently proposed by a few authors, on different grounds and in different ways (Miró-Quesada 1951, 1972, Leonard 1959, Bunge 1973). The particular translation we advocate involves the explicit introduction of values, rights, or duties. For example, we propositionalize the imperative ‘Enjoy life and help live’ as “It is right to enjoy life and help live”. Table 9.1 exhibits a dictionary for translating imperatives into declaratives.

(Warning: Do not give in to the temptation of formalizing prohibitions as $D\bar{x}$ and exemptions as $R\bar{x}$, where ‘$\bar{x}$’ would stand for a non-action, such as not-kicking, because there are no negative facts. This is just a reminder that ethics is not to be divorced from ontology.)

The propositionalization of norms allows us to investigate the logical

TABLE 9.1. Translation of permissions, obligations, prohibitions and exemptions in terms of rights and duties.

Type	Imperative form	Declarative form	Symbol
Permission	You may do *x*.	You have the right to do *x*.	Rx
Obligation	Do *x*!	You have the duty to do *x*.	Dx
Prohibition	Don’t do *x*!	You have no right to do *x*.	$\neg Rx$
Exemption	You are not obliged to do *x*.	You have no duty to do *x*.	$\neg Dx$

relations among them. For example, in order to help someone live one must be alive to begin with; and if one is alive it is because, no matter how hard one's life may be, one does enjoy parts of it. In short, enjoying life, however minimally, is a necessary condition for helping others live. In symbols, $H \Rightarrow E$, i.e. the norm "It is right to help live" implies the norm "It is right to enjoy life". Another example: We take it that the optimization of welfare and the maximization of justice imply one another, i.e. $J \Leftrightarrow W$. Indeed, a person is not dealt with justly unless he is allowed to enjoy some well-being, and in turn he cannot be wholly well unless he is treated justly.

The ability to establish logical relations among norms, via their corresponding propositions, makes it possible to analyze normative systems, such as moral and legal codes, with the help of ordinary logic. A *normative system* is a set of norms, descriptive propositions, and definitions, that (*a*) shares some referents (e.g. human beings or actions of certain types); (*b*) can be disjoined, conjoined, instantiated and generalized to yield further norms in the system, (*c*) contains (potentially) all of the logical consequences of any one or more norms in the system, and (*d*) can be enriched, without contradiction, with further non-normative formulas, in particular data and natural or social laws — i.e. it is an incomplete system. (For an alternative definition see Alchourrón & Bulygin 1971.)

A normative system is structurally (syntactically) and semantically similar to a scientific theory, except that its principles are prescriptive and (more or less efficient) in addition to being descriptive (of ideal types) and (more or less) true. The systems of the two types are also epistemologically similar in that, in order to be applied to a problem or case, the system must be enriched with empirical data and subsidiary assumptions concerning the circumstances of the case or problem. For example, in order to ascertain whether a certain individual has the duty to support his parents, we must find out whether or not he has or could obtain a gainful occupation, and whether his parents are actually destitute. More on this in Sects. 2.3 and 2.4.

(Ideally, the union of a normative system N, a set H of subsidiary hypotheses, and another set D of empirical data, entails the set S of solutions to the problems that motivated the construction of the system N. In short, $N \cup H \cup D \vdash S$. When this proves not to be the case, N, H, D, or all three are either incomplete, wrong, or both — and it is no easy matter to find out which set of premises is at fault. In short, from an instrumental or pragmatic viewpoint, a normative system, be it a technical operating manual, a legal code, or a moral one, is a device for cranking out solutions to practical problems with the help of subsidiary hypotheses and empirical data.)

Legal and moral philosophers are often under the illusion that there can be complete and even final normative systems, i.e. systems that can help solve all deontic problems of a kind. This belief is illusory because normative systems are designed to cope with human actions, and these as well as our knowledge of them vary in kind in the course of time. For example, new criminal methods, such as computer fraud, keep appearing; some practices are criminalized whereas others are decriminalized; and psychology and the social sciences continue to learn about human behavior. What we can reasonably hope for is not completeness but comprehensiveness, the possibility of closing gaps without altering the frame or constitutional principles, and substantial reforms once in a while to cope with radically new social structures and individual aspirations.

1.5 *Moral Reasoning*

Emotivists and intuitionists claim that ethics has little if anything to do with logic. At the other extreme some utilitarians claim to have invented calculi for making correct moral decisions, and deontic logicians maintain that they hold the key to valid moral argument. For our part we believe to have shown in the preceding section that deontic logic is useless, and in Ch. 3, Sect. 3.4 that there is no universal value calculus, whence there can be no universal moral calculus either.

However, the failure of these worthy programs does not entail that we must proceed either irrationally or empirically in moral matters. These are too important to be taken lightly. In moral philosophy, just as in theoretical physics, chemistry, biology, psychology, or sociology, we solve problems by reasoning from a set of premises. Any such set of premises contains general statements (laws or rules), subsidiary hypotheses, and empirical data concerning the specifics of the particular case in hand. Valid proof in ethics does not differ from valid proof elsewhere. In order to prove anything we must start by assuming something else, if only for the sake of argument. In particular we must assume first principles or logical consequences thereof. Such principles will of course differ from one ethical doctrine to the next. Hence, whereas in some contexts (theories) we may prove certain ethical statements, in others we may not: here as elsewhere provability is contextual, i.e. dependent on the assumptions. For example, we can prove that there is a basic human right to work, from the right to life postulate and the general statements that people must eat in order to stay alive, and that food is a product of work (honest or otherwise).

In ethics, like anywhere else, laws (or norms) are of two kinds: basic (or postulates) and derived (or theorems). As in science, in moral philosophy laws (or norms and metanorms) are either universal or particular: either they refer to everyone or only to people in a certain category. By definition, the most basic moral laws, i.e. principles, are universal. For example, "We should make allowances (compensate) for any initial disadvantage or handicap" is a universal law. On the other hand "The blind should be offered assistance in the street" is a particular norm.

Another important distinction is that between unconditional (or categorical) and conditional (or hypothetical) principles. Our maximal moral principle, "It is right to enjoy life and help live", is the only categorical or unconditional one in our system. It suffers no *ifs* or *buts* and offers no reasons or causes: it is the yardstick or standard we use to evaluate morally every action and every other norm. This is because in a humanistic perspective like ours one lives only once and for life's sake — one's own and others's, not for some ulterior motive such as a happy afterlife or the greater glory of God, the nation, the church, or the party.

Two scientific analogs of our categorical principle are a physical standard, such as the standard length unit, and the hypothesis "All electrons are electrically charged". The former serves to measure or calculate, in the last analysis, any length value. As for the hypothesis about electrons, nobody attempts to explain it. Rather, it is so basic that it explains other facts, e.g. that electrons are surrounded by electromagnetic fields and that they interact with other electrically charged entities. Likewise our basic moral norm validates or justifies basic rights, such as the rights to security and work, and basic duties, such as those to protect the helpless and do our jobs the best we can.

Just as in other fields of inquiry, what is a postulate in one system may become a theorem in another. We may derive some norms either directly or via value judgments. In the first paragraph of this section we derived the right to work from three premises. To see how moral laws or norms can be obtained from other general statements together with value judgments, consider the following intuitive argument.

> If *A* (is the case, is done), then *B* (results).
> Now, the result *B* is good (or right).
> Therefore, *A* is good (or right).

This is not a logical but, as we shall say, an axiological inference pattern. It is consequentialist and, as it stands, utilitarian, for ignoring the possible side effects of both *A* and *B*. We account for these side effects by replacing "*B* is good (or right)" with "*B* is good (or right) and, on balance, better (or more right) than *A*".

Even after having propositionalized all imperatives, ordinary logic is not sufficient to account for moral and ethical reasonings. In fact it must be enriched with *axiological inference rules*, the basic ones of which are as follows (modified from Bunge 1973).

Modus volens	
Law	If *A* (is the case, or is done), then *B* (results).
Value judgment	*B* is good (or right) and, on balance, better (or more right) than *A*.
Norm	:: *A* is good (or right) [i.e. *A* ought to be (the case, or done)].
Modus nolens	or
Law	If *A* (is the case, or is done), then *B* (results).
Value judgment	*B* is bad (or wrong).
Norm	:: *A* is bad (or wrong) [i.e. *A* ought to be avoided or refrained from].

Note that, whereas the modus nolens is analogous to the modus tollens of elementary logic, the modus volens is analogous to a common logical fallacy, not to the modus ponens. Moreover, if $A \Rightarrow B$ and *A* is good, but *B* is not evaluated, no course of action can be recommended.

We submit that this approach to the construction of derivative norms is interesting for several reasons. Firstly, it breaks the much-touted fact-value and is-ought barriers. In fact, it combines laws, in particular scientific laws, with value judgements, to produce norms. Secondly, it validates uncounted inferences that we make in everyday life but are not covered by ordinary logic. Thirdly, it strengthens the bond between ethics and axiology. Fourthly, it makes no use of deontic logic. Rather on the contrary, instead of taking all norms for granted as objects of analysis, it invites us to examine their scientific and axiological roots if any, thus favoring a critical approach to morals and ethics. (For the "practical syllogism" approach to the problem see von Wright 1971 and Tuomela 1977, 1984.)

The role of fiction is just as important in ethical reasoning as it is in the arts, technology, or science. And, just as elsewhere, it must be kept under tight control. In particular, we must distinguish between fictitious situations and fictitious entities. Fictitious *situations* may serve heuristic or didactic purposes. For example, we may ask what would happen if everyone were to adopt the immoral mores of the Wild West, or if nobody cared about environmental protection.

On the other hand fictitious *entities*, such as "natural man", "economic man", the "ideal observer", and the "veil of ignorance", do not seem to serve any useful purpose. For example, the so-called ideal observer (or impartial arbitrator) "theory" states that a morally right action is one that *would* be approved of by an omniscient, dispassionate and disinterested person, i.e. a benevolent semigod. Since no such being exists, the proposed definition is useless. Moreover, if an ideal observer did exist, he would lack moral feelings, so that he would be as little justified to judge moral conduct as a bachelor is to advise on marital problems. Rawls's veil of ignorance, allegedly a device for hitting on a just and rational distribution of goods, is equally useless. No such distribution can be designed unless we know something about the basic needs and legitimate wants of the people concerned. As the Stoics emphasized, it takes knowledge to practice virtue.

Parfit (1984) has argued that in moral matters we must focus on what *happens* to people not on *who* they are, whence the concept of a person would be dispensable in ethics. Although it is true that moral principles must be suprapersonal or universal, it is plain that they are about persons, so that morals without persons would be as ghostly as knowledge without knowing subjects. When formulating *general* moral principles we must not try to favor or harm particular individuals. But when formulating *special* moral norms we must take special kinds of persons into consideration, for otherwise there would be no such special norms.

Besides, when *applying* moral principles, whether general or special, a knowledge of personal circumstances does matter. It matters so much that we are forced to employ arguments *ad hominem*, which are intolerable in the sciences. Thus we are justified in being merciful with a destitute and ignorant sinner, and merciless with a wealthy and educated one. Certainly the seriousness of a crime is proportional to the size of the damage, but blame should be inversely proportional to the ability of the offender to lead a clean life. Likewise the praise that a good action deserves is proportional to the size of its outcome and inversely proportional to the power of its actor. It is easy to behave philanthropically when one is well-heeled, hard when barefooted.

Finally, let us remember that logic provides structure not substance. Hence the moral philosopher who is more interested in the former than in the latter may lose sight of moral and ethical principles. Thus, he may examine the arguments used to condemn torture and, finding them logically wanting, he may conclude that torture is justifiable. Unless we care for morals, our metaethics may be an empty shell or worse.

1.6 *Summary*

Ethics may be cultivated as a branch of science or as a philosophical discipline. In the former case the student focuses on the moral norms prevailing in certain social groups, in the latter he pays special attention to the universal or cross-cultural moral norms. If scientific, ethics makes use of the scientific method; if philosophical, it is eminently conceptual. However, if sound, scientific ethics makes use of some of the results of philosophical ethics, and the latter utilizes some of the findings of scientific ethics. Therefore it makes sense to take the union of scientific and philosophical ethics and call it *ethics* without further ado.

In our system ethics is based on both science and value theory. The reason is that ethics deals with right actions, which are defined in terms of actions that increase the well-being of someone, a change that ought to be evaluated scientifically.

Every moral code is constrained not only by nature and society but also by a general frame or moral constitution, as well as by a set of metanorms. The moral constitution of our ethical theory is humanistic rather than theistic, and systemic rather than either individualist or collectivist. As for metanorms, they are conditions on moral norms. One of them is the universality condition, which must be demanded of the basic moral principles. Given the variety of human biological and social groups, a realistic ethics will admit the need for a variety of moral codes. But, since all such codes concern human beings, they are bound to have a non-empty intersection.

Moral codes are normative systems, i.e. systems composed of interrelated norms. In the old times norms were construed as orders — 'do this', 'don't do that'. The general decline of religion and of authority based on brute force has made us more critical. We prefer sentences to imperatives. Fortunately every imperative can be translated into a sentence — though of course not conversely. For example, 'Do x!' translates into 'You have the duty to do x', and 'Don't do x' into 'You have no right to do x'. Unlike orders, sentences about rights and duties can be argued about, and some of them can even be tested. The exception is the maximal norm "Enjoy life and help live", which — when suitably propositionalized in terms of rights and duties — functions as *ultima ratio.*

Reasoning is no less important in ethics than in any other field of knowledge. By and large it fits standard deductive logic. However, any consequentialist ethical system, like ours, will require, in addition, certain patterns of axiological inference leading from the conjunction of laws and value judgments to (derivative) normative statements. We have singled out two of them, which we have called *modus volens* and *modus nolens.* Such (nondeductive) derivation of norms from laws and values suggests that the famous is-ought gap is not an abyss after all. But this is a matter for the next section.

2. ETHICS AND REALITY

2.1 *The Is-Ought Gap*

Most philosophers accept Hume's thesis that ought-statements cannot be logically derived from is-statements: that there is an unbridgeable chasm between moral norms and factual propositions, hence between ethics and science. However, a few philosophers have challenged Hume on this point. Let us take a quick look at some of their arguments.

Black (1964) claimed to have dismantled what he called 'Hume's guillotine' by means of such alleged counterexamples as this:

> You want to achieve B.
> Doing A is the one and only way to achieve B.
> Therefore, you should do A.

But it is evident that the conclusion does not follow *logically* from the premises, for these fail to contain the "should" (or "ought") concept. The above "practical syllogism" is an instance of what we have called an axiological inference pattern, namely the modus volens (Sect. 1.5). This rule of inference is reasonable, even necessary in ethics and technology, but it does not fit deductive logic.

Mackie (1977) held the following to be a valid inference: "Doing *X* is wrong, therefore you ought not to do *X*". This looks like a valid inference only because we tacitly assume that one ought not to do anything wrong; but as it stands it is not valid. If the missing premise is explicitly added, we do obtain a valid inference, namely

> You ought not to do anything wrong.
> Doing X is wrong.
> Therefore, you ought not to do X.

(If in doubt about the validity of this argument, translate the major premise into "You have no right to do anything wrong", and the conclusion into "You have no right to do *X*".)

MacIntyre (1984) maintains that the following example refutes Hume. From the factual premises that this watch is inaccurate, irregular, and too heavy, the value judgment follows that this is a bad watch. Actually nothing of the sort follows: MacIntyre's is an enthymeme not a deductive argument. In order for the conclusion to follow we must add either an explicit definition of "bad watch" or the conditional "If a watch is inaccurate, irregular, or too heavy, it is a bad watch". In conclusion, *Hume ab aevo vindicatus.*

We admit the is-ought gap but deny that it is a chasm, and submit that we cross it daily both conceptually and practically. We do the former when constructing "practical syllogisms" such as the *modus volens* and the *modus nolens* introduced in Sect. 1.5. In more complex cases we may need not only value judgments but also higher level moral norms in order to derive the desired lower level moral norm. For example, to derive the norm "Protect the environment", or better "If you value life protect the environment", we need to know not only that normal life requires a clean enviornment: we also need the moral norm "Enjoy life and help live". Another example: Whenever we evaluate possible courses of action, we pick the one that seems best, and proceed to implement it, on the strength of the moral principle *We ought always to do the best* (or *We have the duty to do our best*). In short, we arrive at norms either from value judgments conjoined with factual statements, or from such premises enriched with higher level norms.

As for the practical bridging of the is-ought gap, it consists in acting in order to meet needs or wants: in transforming a deficient *is* (current state of affairs) into a satisfying *ought* (future state of some system of which the agent is a part). This transition can be described in cybernetic terms within the state space schema adopted in Vol. 3, Ch. 3, and Vol. 4, Ch. 1. In fact we may proceed as follows.

Saying that agent X *ought* to do G amounts to saying that G is a *goal state* (or set of states) of X, or of some system Y under the control of X, and that G is more valuable to X than the present state of X or of Y, as the case may be. The goal state G is specified and achieved either automatically (as in biological control) or deliberately (as when reaching for a banana or turning the knob of a thermostat). In either case the intensity of the drive to attain G is an increasing function of the discrepancy between G and the current value $F(t)$ of the state function of the system under control. When the organism or the system it controls is equipped with a feedback mechanism, it is capable — barring strong unforeseen perturbations — of attaining G by successive approximations (e.g. oscillations of decreasing amplitude). In normal cases the goal coincides with the objectively good or right, so that the transition from the current state to the goal state improves the well-being of the organism. In abnormal cases (e.g. self-mutilation and trading bread for guns) the system goes to a worse state. In either case, whether G is good (or right) or bad (or wrong), every time the system attains G, the *ought* becomes *is*, i.e. Hume's gap is being closed. See Figure 9.5.

2.2 *Moral Truth*

Most philosophers deny that there can be moral truths because they conceive of moral principles as orders expressed by imperatives, and obviously these can be relevant or irrelevant, efficient or inefficient, but not false or true. This objection is invalid if it be admitted that imperatives are translatable into declaratives, as proposed in Sect. 1.4. For example, "Killing is wrong" is true in any ethics asserting the right to life, false in any alternative ethics.

The truth or falsity of a moral principle can best be investigated when spelled out in terms of rights or duties, of virtues or sins, or of good or bad consequences. For example, "Help thy neighbor" can be

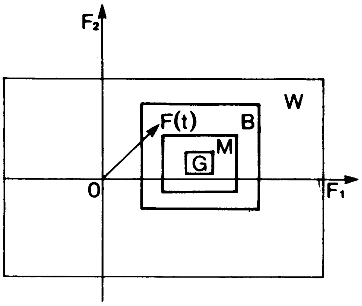


Fig. 9.5. Closing Hume's gap in practice. The system is in state $F(t) = \langle F_1(t), F_2(t)\rangle$ at time t, and it tends to go into one of the good states lying within the rectangle G. The states outside G are mediocre (M), bad (B), or wretched (W). The system does not work at all as such if in a wretched state. The transition from a B or an M state to a G state closes the is-ought gap.

propositionalized as "It is right to help one's neighbor". And the truth of this precept becomes obvious when spelled ou as "If you care for your neighbor, or wish to be on good terms with him, or be able to count on him in an emergency, or be at peace with your conscience, help him". Likewise "Thrift is virtuous and waste sinful" is true in a world of increasing scarcity, even though it will not meet with the approval of short-sighted economists and statesmen.

As with moral truth, so with falsity. For example, "What is good for the economy is good for the individual" is false, because the economy may be stimulated in the short run by wasteful consumption, which in the long run is bad for humankind. (See Veblen 1899 p. 73.) Likewise "What is good for the state (or GM or the church or the party) is good for all" is false whenever the state (or corporation or church or party) hardly cares for the individual well-being of everyone.

Since there are moral truths there can also be moral lies, i.e. utterances that deliberately convey moral falsities. For example, the claim that there are inferior races is not only a scientific falsity: it is also a moral lie used to justify the exploitation or oppression of certain ethnic groups. And, whereas the utterance of a factual falsity may be merely a mistake, that of a moral lie is almost always sinful. The exception is of course the white or merciful lie.

The thesis that there are moral truths and falsities presupposes that there are *moral facts*. (In turn, this presupposes that moral truth is not formal, and that factual truth is correspondence with fact.) The thesis that there are moral facts is denied by most philosophers. (Harman 1977 and Nino 1985 are exceptions.) But in Sect. 1.1 we argued that there *are* facts of this kind, namely all those involving rights or duties. For example, dumping toxic waste in a lake is a moral (or rather immoral) fact because it violates the right of others to enjoy a clean environment, (Consequently the proposition "Dumping toxic waste in a lake is wrong" is a moral truth.) Seeking a cease-fire in a war, and signing a non-aggression pact, are moral facts as well for helping exercise the right to life. Consequently the proposition "Peace-making is right" is a moral truth — as is "Warmongering is wrong".

However, there is a difference between moral truths and physical or biological truths, namely that the former are *contextual* whereas the latter are not. Indeed, since moral truths concern rights and duties, and these are relative to the moral code, the former are contextual as well. In this regard, and only in this one, moral truths are akin to mathematical truths, every one of which holds within a given system (theory) but may be false or even meaningless in alternative systems. But even this context-dependence is limited, for all viable moral codes share some principles.

The natural law theorist claims of course that moral rights and duties are absolute, i.e. context-free. But since, as a matter of fact, certain rights and duties fail to be enshrined in a number of moral codes, he cannot claim that the corresponding propositions are true. If he nevertheless does make this claim, then he cannot hold such truths to correspond to facts: he must hold them to be formal like mathematical theorems, or he must adopt a non-realist concept of truth. Given these difficulties he is likely to retreat to the traditional view that moral norms can be neither true nor false. But then he cuts the root of his doctrine to nature and natural science, so he has no right to call himself a *natural* theorist. The aim of this discussion has been to support our thesis in Ch. 8, Sect. 1.2, that ethics cannot be divorced from semantics and epistemology.

In conclusion, there are moral truths and falsities because there are moral facts. Consequently moral principles are testable. More on this in Sect. 2.4.

2.3 *Applying Ethics*

Does ethics make a difference to real life, i.e. are ethical doctrines ever applied with any success? At first sight the empirical evidence is equivocal. For one thing some experiments in social psychology seem to show that, in similar circumstances, people behave similarly regardless of their "ethical ideology" (Forsyth & Berger 1982). The historical record would seem to support this hypothesis. For example, we all know that Christian ethics has inspired as many crimes as selfless deeds, and that utilitarianism has inspired progressive social reforms as well as egoistic behavior.

Furthermore, much as philosophers would like to believe that Spinoza inspired Locke, who in turn inspired the French Enlightenment, which begat the French Revolution, which enshrined the noble devise *Liberté, égalité, fraternité*, historians know better. They know that, at best, ethical ideas *facilitate* social movements. (Incidentally, the thesis that Protestant ethics begat capitalism, holds no water. In fact (*a*) capitalism had existed on a small scale before modern times, and (*b*) modern capitalism was born more or less simultaneously in Catholic as well as in Protestant countries. What is true is that Protestantism *favored* capitalism more than Catholicism did. Weber [1904—1905] wavered between the two theses.)

Most people go by certain moral norms, such as those of equity and reciprocity, that are not the prerogative of any single moral code or ethical doctrine. (For example, the Golden Rule occurs in most moral codes, and most ethical theories concur with the thesis that morality has a social function.) In other words, it is not that ethics makes no difference, but rather that there is a certain minimal morality, and a certain minimal ethical doctrine, that are universal or nearly so. Whenever we confront complex moral issues we make explicit use of some such norms. And whenever we do so we face the methodological problem of solving individual cases in the light of general principles.

According to a popular view, inherent in legal positivism, the application of a rule to a particular case is straightforward: the rule either fits the case or it does not. People with real life experience, particularly in moral, legal or political matters, know better: they know that the counselor, judge or politician must "interpret" the rule, and that such an "interpretation" is not really a semantic affair but that it consists in the addition of data and subsidiary assumptions. Codes and theories, whether moral, legal, scientific or technological, are only *sketches* that must be supplemented with particulars if they are to be applied to individual cases. Such particulars concern not only the case under examination but also precedents about similar cases in the past. Such precedents are among the data to be used to form an opinion or reach a decision on the case in hand. (For the logic of such decision see Sect. 1.4.) However, in moral, legal and political matters decisions do not follow logically from principles, hypotheses and data. For better or for worse, normative reasonings are always contaminated or purified by feeling and bias.

Furthermore, no moral or legal principle can be applied singly without harm, for every such norm focuses on a limited area of human action, whence it leaves other areas in the dark. If we seek moral or legal justice, every time we are called upon to apply a moral or legal principle we must bring to bear its companions as well. In short, we must proceed *systemically* — though not holistically. (See Vol. 4.) When doing so we realize that no two moral or legal problems are identical. Thus if person *A*, a college graduate with a good job, and *B*, a high-school drop-out without a job, commit the same (moral or legal) offense, it would be unfair to give them the same sentence. *A* deserves a harsher punishment than *B*, for *A* ought to know better and he can meet his basic needs and legitimate aspirations without breaking the law, whereas *B* cannot. (In real life things often happen the other way round: *A* can hire a better advocate than *B*, so he is likely to get off with a lighter sentence, particularly if *A* is a member of the dominant ethnic group and *B* is not.)

An extremely demanding and rigorous moral code or ethical doctrine is unlikely to find lasting application in the modern world. To be viable a moral code or an ethical theory must be (*a*) *realistic*, i.e. account for the basic needs and legitimate aspirations of people; (*b*) *socially useful*, i.e. inspire prosocial behavior and discourage antisocial conduct; (*c*) *flexible*, i.e. adaptable to new circumstances; (*d*) *compatible* with the best available knowledge about human nature and human society, and (*e*) *equitable*, i.e. help to compensate for inequalities and erradicate privilege. In short, the truer and the more right, the better its chances of being applied with success rather than remaining buried in books.

2.4 *Testing Ethics*

In Sect. 1.1 we distinguished scientific from philosophical ethics. The hypotheses and theories in the former are, by definition of 'scientific', just as testable as those in any other scientific disciplines. The question is whether philosophical ethics is testable and, if so, how. For example, how do we go about finding out whether moral nihilism or egoism, conventionalism or emotivism, is true (to fact) and efficient (in guiding action)? Emotivists, intuitionists, natural law theorists, deontologists, conventionalists, moral positivists and other ethical dogmatists do not face the test problem. On the other hand those of us who believe that there are moral facts and moral truths, that some moral norms can become obsolete or even obnoxious, and that moral principles should help govern efficient action, do face the test problem.

In our view there is no major problem with testing ethical *theories*, i.e. hypothetico-deductive systems about the nature, root and function of moral norms: such theories are testable the same way as psychological or sociological theories, namely by their consistency with both facts and other theories. (See Vol. 6, Ch. 12, Sect. 3 for the need for such a dual test.) For example, we rejected ethical emotivism and intuitionism for failing to account for the rational and empirical inputs to moral deliberation; and we rejected utilitarianism for overlooking moral sentiments and for using subjective values and probabilities. (Recall Ch. 7.) From an ethical theory we demand (*a*) internal logical consistency; (*b*) external consistency, i.e. compatibility with the bulk of relevant scientific knowledge; (*d*) ability to account for viable moral codes; (*d*) helpfulness in suggesting needed moral reforms, and (*e*) usefulness in analyzing moral concepts and principles, hence (indirectly) in settling moral disputes. The testing of ethical theories is thus similar to that of theories about grammars — for which see Vol. 7, Part II, Ch. 4, Sect. 3.2.

What does pose a serious methodological problem is the test of any moral *norm*, i.e. the justification of the claims that the norm is (*a*) a moral *truth* (or falsity) for fitting (or failing to fit) certain moral facts, and (*b*) practically *efficient* (or inefficient) for promoting good (or bad) moral conduct. In our view all moral norms except the highest of them all — "Enjoy life and help live" — are testable provided they are stated in clear terms. They are testable in three different but mutually complementary ways: (*a*) by checking their consistency with higher moral or ethical principles; (*b*) by checking their consistency with the relevant scientific, technological or ordinary knowledge pieces of information, and (*c*) by ascertaining whether they affectively contribute to maintaining or enhancing individual or social welfare. Let us examine these three types of test.

Moral norms, just like scientific hypotheses, must be *consistent with the highest accepted principles*, in this case the first moral and metaethical principles in the system concerned. In the case of our system, the highest principle to which every other moral rule or meta-rule must conform is of course "Enjoy life and help live". Any rule that facilitates the implementation of this principle is morally right, and any rule that hinders it is morally wrong. Example of the former: "Know thyself and thy neighbor". Example of the latter: Any rule that enforces discrimination on any grounds other than ability.

Einstein (1950 p. vii) showed clearly how a (derivative) moral norm can be validated by "tracing it back" to basic moral precepts. The injunction not to lie can be justified as follows. "Lying destroys confidence in the statements of other people. Without such confidence, social cooperation is made impossible or at least difficult. Such cooperation, however, is essential in order to make human life possible and tolerable. This means that the rule 'Thou shalt not lie' has been traced back to the demands: 'Human life shall be preserved', and 'Pain and sorrow shall be lessened as much as possible'". In other words, we are returned to our supreme maxim "Enjoy life and help live". (Interestingly enough the great theologian Suárez (1612) advocated the same method. In fact he held that the validity of a norm derives ultimately from the "eternal laws" and the "natural laws": *Lex aeterna* implies *lex naturalis*, which in turn implies *lex positiva*.)

The second test is that of *consistency with the best available knowledge*. Thus a moral norm that were to ignore moral sentiments, or were at variance with the well known sociological finding that moral rules are human creations that affect (positively or negatively) human relations, could be discarded out of hand. For example, a norm that commands making human sacrifices to a divinity is not only cruel but also incompatible with anthropology and history, both of which show that divinities are human inventions. Likewise, social science teaches us that certain food and sexual taboos, ordinarily thought to be moral norms, are nothing but customs — albeit often customs with a social or religious rationale. (See e.g. Harris 1979.)

The third and last test to which any moral norm ought to be submitted is that of *efficiency*. It consists in checking whether the norm in question helps realize the underlying basic values. In this regard moral norms are similar to the science-based rules of modern technology: both are based on laws and they are tested by their efficiency in helping bring about desirable states of affair. And, like the efficiency of technological rules, that of moral norms is best assessed by scientists and sociotechnologists engaged in studying people. Moral codes are right if they "work", wrong if they don't. But in order to work they must be viable to begin with, and such viability calls for realism (Sect. 2.4).

Even if a moral norm passes all three tests, it must be regarded as imperfect. Four centuries of scientific and technological experience, not to mention philosophical criticism and social experience, should prevent anyone from pursuing infallibilist mirages, in particular that of an *ethica perennis*. We all should know by now that no moral code and no ethical theory can guarantee anything, even though some codes and some theories can be more helpful than others. In short, we should be ethical *fallibilists*, i.e. ready to admit moral or ethical error, rather than being dogmatists. (For the dogmatic feature of moral intuitionism see Ch. 7, Sect. 1.5.)

But fallibilism is destructive unless accompanied by *meliorism*. We may and should hope for moral and ethical progress through both social reform and theoretical work. Ethical meliorism is justifiable on the following grounds. Firstly, morals coevolve with society, and ethics with the rest of philosophy as well as with social science and social technology. Secondly, if we want to we can spot and correct moral deviations and ethical mistakes, though it may take some intellectual or even civic courage to do so.

Take for example the slogan *Freedom or death*, embraced by countless political leaders since ancient times. At first blush every freedom lover should adopt it. But on second thoughts it becomes obvious that the precept is optional, not binding on everyone. Indeed (*a*) even the life of a slave may be worth living; (*b*) sometimes the bondman has a real possibility of gaining freedom by peaceful means; (*c*) nobody has the right to bully others into a fight to the death; and (*d*) the price of any war without quarter can be staggering, not only in terms of casualties but also because it makes it hard for the defeated, whatever the outcome, to retain or to gain any freedoms. There are only two cases where fighting to the death or committing suicide is preferable to servitude: when faced with a merciless enemy and when the price of freedom is betrayal. In all other cases we must think twice before resorting to violence in order to conquer freedom, and this for prudential as well as for moral reasons. (Recall Ch. 8, Sect. 2.5.) In any event, the point was to suggest that once in a while we must question our most cherished moral beliefs, for fresh information or new thinking may disprove them.

In conclusion, "Ethical axioms are found and tested not very differently from the axioms of science" (Einstein 1950 p. viii). Both are ultimately vindicated or invalidated by experience. In the cases of moral norms and ethical theories the empirical data concern human welfare. Hence biological and social indicators — such as life expectancy and infant mortality, protein intake and average number of school years — are more relevant to the testing of moral norms and ethical theories than pondering over interpersonal relations or exceptional moral predicaments (Waddington 1961 p. 18).

2.5 *Summary*

There is a gap between that there is and what there ought to be, but this gap can be bridged. We cross it daily both conceptually and practically. We do this by thinking that we ought always to do what is best and by attempting to do it. Meeting ends and wants, i.e. transforming what ought to be into actuality, is what life is all about.

Since any imperative can be translated into a declarative, moral norms take the form of propositions, in particular conditionals, whence they are true or false to some extent. For example, "Keep your promises" translates into "It is right to keep one's promises". This is a true proposition, for breaking promises causes harm, hence it is wrong. Moral truths match moral facts. Moral falsities, in particular moral lies, distort reality and may put welfare at risk.

Moral norms regulate some of our actions; and occasionally we design policies and plans in the light of ethical doctrines. The application of either is at least as complex as the application of scientific hypotheses and theories. Far from being straightforward, it involves the gathering of data concerning circumstances and precedents. Moreover, moral rules ought never to be applied singly because every moral problem is complex. Here as elsewhere systemic thinking is the ticket.

Moral norms and ethical theories are to be tested in two ways: for truth and for efficiency. Ethical theories are tested for truth much in the same way scientific theories are; their test is particularly similar to that of the test of theories about grammars. As for moral norms, the maximal one is not directly testable. But its consequences are testable and it is a yardstick for evaluating all of the other norms. The latter are tested in three ways: by checking their consistency with the highest level norms, by seeing whether they harmonize with the best relevant knowledge, and by observing in practice whether or not they contribute to overall welfare.

But even if a moral norm passes all three tests it must be regarded as provisional, either because it is based on defective knowledge or because it may become irrelevant. That is, we must espouse ethical fallibilism, though not to the point of radical skepticism. Here as elsewhere we are justified in holding on to a bunch of firm principles, such as that of mutual help, as well as in believing in the possibility of improvement. In short, we are meliorists as well as fallibilists.

We have come to the end of Part III on ethics. The results obtained so far in the book will be employed to study human action, particularly of the social kind, in the next part.

PART IV

ACTION THEORY

CHAPTER 10

ACTION

So far we have studied the good and the right. In the following we shall inquire into the sorts of action that we should perform in order to realize our values in a practically efficient and morally right way. That is, we shall sketch an action theory and a social philosophy. These ideas will be applied in the last chapter to discussing the most pressing issue of our time: the survival of our species.

Human action is a subject of study of scientists, particularly sociologists, and philosophers. The former are mainly interested in collective or social action, whereas the latter focus their attention on individual action. Typically, social scientists attempt to explain social action not only in terms of individual intentions but also in terms of social forces, whereas philosophers are only interested in intentions and decisions. Most social scientists do not care to elucidate the key concepts they employ in describing or explaining action, whereas most philosophers are exclusively interested in such elucidation. (For the contrast between the two approaches see Seebass & Tuomela Eds. 1985.)

We submit that there need be no contrast between the scientific and the philosophic approaches to human action and, moreover, that it is convenient to combine the two approaches. In particular, it is possible and desirable to analyze social action into a collection of individual actions (under social constraints) with similar purposes (though perhaps different motivations). But such analysis of social action in individualistic terms must be supplemented with an examination of the social outcomes of concerted individual action. That is, social action does not affect solely the personal lives of the actors involved in it: it may also alter the structure of the society, to the point of creating or destroying entire social groups or institutions. Typically, the philosophical analyses of social action disregard such social changes for ignoring large-scale social movements such as strikes, wars, and the building of business or political empires. If only for this reason, i.e. to shift from trivia to genuine problems, it is healthy for the social philosopher to rub shoulders with social scientists. What holds for action theory holds for social philosophy as well.

The orientation of our action theory and our social philosophy is scientistic rather than aprioristic or speculative. (For an examination of scientism see Vol. 6, Ch. 15, Sect. 2.2.) Consequently we do not regard action theory as an exercise in logic, or social philosophy as a system of dogmas concocted without the help of social science and sociotechnology. In particular, we have no use for modal logic or for the metaphysics of possible worlds. We claim that a realistic theory of action uses the concept of real possibility, which can only be elucidated in terms of objective regularities (laws and rules), constraints, and circumstances. We also claim that, if we are to understand human action and help design it, we must look at the real world rather than imagine possible worlds.

Taking a scientistic approach to human action and social issues does not involve engaging in a value-free, ethically noncommittal, or ideologically neutral speculation. Far from this, our action theory and our social philosophy depend on the value system and the moral norms adopted in previous chapters, and those theories contain ideological theses such as "Integral democracy is the best social order". But we hope that, unlike alternative ideological tenets, ours can be supported by argument and empirical research, and that they serve the interests of humankind rather than those of special groups.

1. INDIVIDUAL ACTION

1.1 *Philosophical Underpinnings*

We start by declaring the philosophical presuppositions of our action theory, in the first place the logical ones. A moment ago we stated that we do not need modal logic, which we regard as a purely formal discipline without applications in science or philosophy. (Vol. 3, Ch. 4, Sect. 1.2.) Nor need we resort to three-valued logic, which has been recommended for allegedly accomodating inaction or abstention as an alternative to approval and disapproval. Ordinary logic suffices to account for this case, as suggested by the following example. "I will join you in doing A if and only if I approve of A. But it so happens that I neither approve nor disapprove of A. Hence I will not join you in doing A." As the reader may easily check, this is merely an instance of the modus tollens for equivalences. Ordinary two-valued logic suffices to handle inaction.

It has also been claimed that action theory needs some theory of contrary to fact conditionals (or counterfactuals). An example of a counterfactual sentence is 'If *A* were to do *B*, then *C* could happen', where it is tacitly being assumed that *A* won't do *B*, perhaps because *A* realizes that the effect *C* is undesirable. There is no doubt that we employ counterfactuals when deliberating about possible courses of action. The question is whether counterfactuals merit a special theory. In our view they do not because every counterfactual may be regarded as a sentence designating either a hypothesis with a false antecedent, or an abbreviated inference with a false premise (Bunge 1973 Ch. 1). Here again ordinary logic is sufficient.

As for the epistemology of action theory, we take sides with realism and against all varieties of anti-realism, in particular subjectivism or relativism, according to which "there is no such thing as the way the world really is". We adopt realism because it is the epistemology inherent in scientific research and in technology (Vol. 6, Ch. 15, Sect. 2.2, and Tuomela 1985). It also underlies any adequate and useful theory of human action. In fact, if there were no objective truths, however partial, we would be unable to use them to plan any efficient courses of action: all our actions would be impulsive. As well, without the notion of objective truth that of factual error would make no sense, and it would play no role in explaining any technical mistakes in the planning of action. Moreover, if all truths were subjective, so would be all evaluations of the extent to which our actions succeed in attaining our goals.

As for the fact that every one of us participates in some social process, rather than being a passive spectator of the social scene, it supports realism rather than subjectivism or relativism. Indeed, if you decide to act upon somebody you presuppose that your prospective patient exists objectively. Furthermore, if you wish your action to be efficient, you had better study the way your prospective patient really is. Many of our failures in dealing with people can be traced back to our tendency to rely on first impressions and prejudices. In short, critical (and in particular scientific) realism is just as crucial in action theory as in science and technology. Anti-realism can only be used as an excuse for inaction: it is nothing but an exercise in academic ingenuity.

We also take sides with the thesis of the basic methodological unity of all the factual sciences, and consequently against the so-called *Verstehen* (or empathic understanding) school. The latter claims that human actions cannot be studied scientifically because some of them are intentional — or, as Weber put it, have a "meaning". This is obvious, but all it proves it that an adequate theory of human action cannot be a branch of physics or chemistry. Physiological psychologists study intentional (or voluntary) acts in the laboratory in the same objective way they study unintentional actions. Moreover, the mere existence of sciences that are social as well as natural — such as geography, demography, social psychology, anthropology, and bioeconomics — refutes the historicist and hermeneutic thesis that the social sciences cannot mix with the natural ones. In short, the historicist (or idealist) philosophy of social science has been left behind by scientific research and therefore it cannot be a sound basis for a theory of human action.

As for the ontology inherent in action theory, let us only make a few brief remarks on two questions, those of determinism and materialism. By admitting the possibility of freedom (Ch. 6, Sect. 2.3) we have rejected classical determinism. However, we have not derived the possibility of freedom from the alleged lawlessness of the mental or from the claim that the basic physical laws are probabilistic rather than causal. We have assumed that both the freedom of choice among given alternatives and the original creation of new alternatives are properties of certain plastic regions of highly evolved brains (Vol. 4, Ch. 4, Sect. 5.1). It is a tacit assumption of every science that every property is related lawfully to some other properties, and this holds in particular for the properties of the human brain. If there were no laws, or if we did not know any of them, we would be unable to act in a deliberate or intentional fashion. Thus when doing A in order to get B, we assume tacitly or explicitly (if only for the purpose of trial) that A is likely to bring B about, either always or with some probability. In short, deliberate human action involves both freedom and lawfulness.

Another ontological problem related to action theory is that of the reality and materiality of the external world. In theory almost any ontology is compatible with human action. Whether by acting we construct the world or merely alter it, is immaterial as long as the desired changes are brought about. But in practice metaphysical idealism is only compatible with quietism, because one does not succeed in altering the world by pure thought. It was not by accident that Buddha taught the unreality of the world and preached inaction, whereas Marx was a materialist of sorts and exalted praxis. The social reformer assumes tacitly that the persons and social groups he deals with exist by themselves, and he endeavors to do something to change them in some way or other. For this reason he must regard metaphysical idealism as impractical or even obstructive: he must assume that the real world is material as well as knowable. (Whether he realizes this is another matter.)

In sum, we seek to sketch an action theory and a social philosophy in harmony with the rest of our philosophical system and, in particular, compatible with our value theory and our ethics, as well as with our materialist ontology and our realist epistemology.

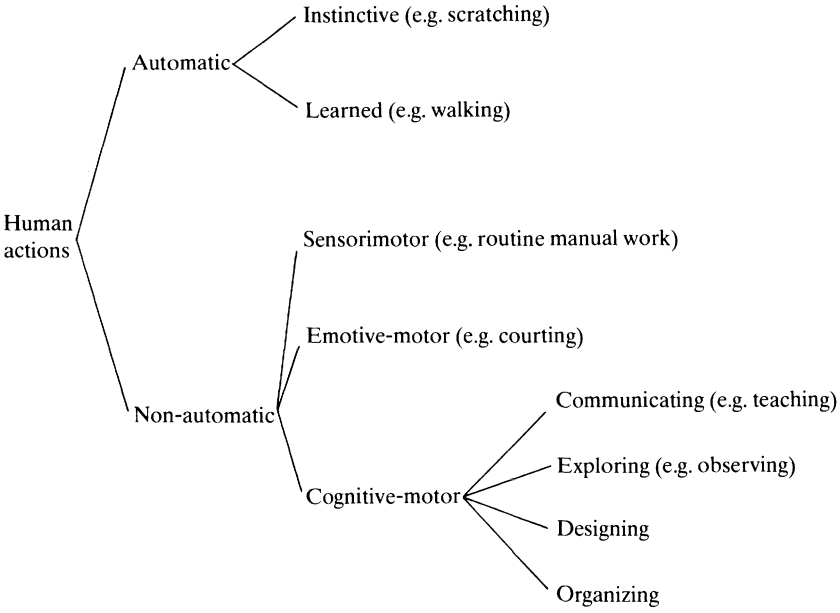
1.2 *Human Action*

All concrete (material) entities are active in some way or other: there are no changeless and totally passive things in the world. Human action is a particular case of action. The action that one thing or its proxy (the agent) exerts upon another thing (the patient) may be defined as the difference that the former makes to the history of the latter. (Recall Vol. 3, Ch. 5, Sect. 4.1.) In human action theory the agent is a human being, or an animal or a machine under his control, and the patient may be any concrete thing, whether human or not, that coexists with the agent at least during part of the period under consideration. There is interaction if the patient reacts upon the agent, as in the cases of work, play, and conversation. And the action is social if both agent and patient are members of the same animal species.

Humans are distinguished from other things not for being doers but for being capable of acting rationally in a morally right or wrong way, i.e. for being able to use knowledge to do good or evil. In short, we are (sometimes) rational moral agents. A nonhuman animal may do another a good turn, but not nearly as rationally, hence effectively, as a human. And an intelligently programmed robot may behave rationally (by proxy) but not with good or evil purposes — except again by proxy — for it is incapable of having intentions of its own. This is not to say that man is the only rational and ethical animal. But, among all the animals known to us, we are certainly the ones capable of acting in the most rational and right way — as well as the only ones capable of putting the highest reason in the service of evil.

The collection of human actions can be partitioned in various ways. We shall make use of the following partition shown overleaf.

Action theory is only concerned with non-automatic human actions. All of them are learned and moreover intentional or goal-directed, even


Human actions
Automatic
Instinctive (e.g. scratching)
Learned (e.g. walking)
Non-automatic
Sensorimotor (e.g. routine manual work)
Emotive-motor (e.g. courting)
Cognitive-motor
Communicating (e.g. teaching)
Exploring (e.g. observing)
Designing
Organizing

though their goal may be only barely sketched rather than designed in detail. Unlike a reflex action, an intentional one is preceded by deliberation and decision, both of which include valuation. Consequently an intentional action is one that we can choose *not* to perform, and one for which we must take full individual responsibility.

Intentional actions may be characterized as positive (or constructive) or as negative (or destructive), depending on whether their goal is to produce or maintain, or else to remove or destroy. Thus hiring an employee is a positive action whereas firing him is a negative one. However, this dichotomy is contextual or situational and, since situations are changeable, it is also dynamic: actions that are positive to some people, at some time, may be negative to other people later on. Thus building one more highway may solve an immediate problem for some motorists, but it may have long term negative consequences for everyone. Life requires taking negative as well as positive actions. However, it is desirable to minimize the former if we are to minimize suffering.

Even a positive or constructive human action has some negative or destructive side, such as waste production and, in general, an increase in the entropy or disorder of the environment. It is therefore advisable to speak of actions that are *mainly*, but never totally, constructive. If we are to live we cannot help increase the environmental entropy or disorder. However, we can avoid wanton destruction, in particular that of highly complex and valuable systems, such as human beings. Since we can and should, we had better state explicitly

NORM 10.1 Do not deliberately increase the overall disorder unless it is for the purpose of increasing the local order in the process of meeting some basic need or legitimate want.

NORM 10.2 Do not destroy anything valuable unless you can replace it with something better without at the same time preventing others from satisfying their basic needs.

These two norms entail (in an intuitive not in a logical sense)

NORM 10.3 Do not design or manufacture artifacts that (*a*) serve no constructive purpose (i.e. are either purely destructive or useless), or (*b*) require the irreversible destruction of a natural resource, or (*c*) have a built-in obsolescence.

This last norm enjoins us not to perform certain actions. Inaction is not a form of action, yet it may have positive or negative consequences, for it allows nature to take its course, or others to take action. This is why morality and the law blame us not only when we do something evil but also when we fail to take action to prevent harm to life or property. By not taking action the onlooker fails to divert the course of events and thus makes no contribution to the prevention of undesirable events. However, in and by itself inaction, not being a process, lacks causal efficacy. (More on this in Hart & Honoré 1959, and in Bunge 1982 b.)

Any course of rational human action may be represented schematically by the flowchart on the next page.

A person behaves rationally if and only if she makes use of the best knowledge available to her in planning her course of action, and it she is willing to discuss her plans and decisions as well as to revise them in the light of new information. (See Mosterín 1978 pp. 56—57.) Acting on impluse, without deliberation and planning, is irrational. So is planning without the benefit of solid knowledge or without making room for unforeseeable disturbances and risks. The same holds for actions which are not monitored or the outcome of which is not evaluated. Yet both in private life and in the public sphere we often do not bother to evaluate some of the most important actions.

Problem
↓
Deliberation 1
↓
Decision 1
↓
Inquiry 1
↓
Knowledge & valuation 1
↓
Belief 1
↓
Decision 2
↓
Plan
↓
Deliberation 2
↓
Decision 3
↓
Inquiry 2
↓
Knowledge & valuation 2
↓
Belief 2
↓
Procurement of means
↓
Organization
↓
Action
↓
Monitoring and control
↓
Outcome
↓
Evaluation of outcome
↓
Cost/benefit and ethical analyses

However, rationality is not enough: the moral agent is supposed to behave rightly as well as rationally. And of course there is room for values and morals in nearly every one of the phases of a course of intentional action, from the spotting of the problem — which, if moral, may not even be "seen" by the moral moron — and deliberating what if anything is to be done about it, to choosing the means to solve the problem, and evaluating the final outcome. Consequently any plan of action may, and ought to be examined from a moral viewpoint as well as from a purely technical one. Technology without morals can be evil, just as morality without technology may be impotent.

Moral philosophers and psychologists are interested in uncovering the possible sources or motivations of human action. It turns out that, whereas some of them are positive or constructive, others are negative or destructive. Basic needs, legitimate interests, love, the sense of duty, and curiosity are among the former; and fear, shame, greed, and hatred among the latter. In a good society the positive motivations would be rewarded and the negative ones discouraged, so that on the whole the former would prevail over the latter. A society that thrives on fear or shame, and promotes greed or hatred, is evil.

Up to a point, the effort put in attaining a goal is the greater, the greater the difference between the value of the goal state and that of the current state of the system that is being acted on. (Physical analog: Ohm's law.) But if the path to the goal proves much longer than foreseen, or the means to attain it do not become available, the agent is justified in reviewing the entire operation. And if the endeavor looks hopeless the rational thing to do is to redesign it *da capo* or even to give it up in time to cut the losses. However, these are practical problems of marginal ethical interest, except of course when they originate in moral problems. The solution to any such practical problem calls for a number of items of specialized knowledge concerning the system of interest. Nevertheless, from such specialized studies a certain number of descriptive and normative principles can be formed. The field of knowledge that seeks to find such generalizations is called *praxiology* and it may be regarded as a branch of general technology and, more particularly, of the general theory of systems. (See Kotarbinski 1965, Gasparski & Pszczolowski Eds. 1983.)

The philosopher is more interested in the justification of actions than in their design, which he gladly leaves in the hands of experts. For one thing he is interested in distinguishing different though related concepts of justification, which we elucidate by means of

DEFINITION 10.1 For any intentional (or deliberate or voluntary) human action A,

(i) A is *conceptually justified* $=_{df}$ A is taken in the light of the body of antecedent knowledge (i.e. A is planned, monitored and evaluated with the help of the best available knowledge relevant to it);

(ii) A is *empirically justified* $=_{df}$ the empirical (observational or experimental) study of actions of the same kind as A has shown that they are efficient or at least effective, and that their desirable effects by far outrank their undesirable side effects;

(iii) A is *practically justified* $=_{df}$ A is both conceptually and empirically justified;

(iv) A is *morally justified* $=_{df}$ There exists at least one moral code according to which A is a right action;

(v) A is *wise* $=_{df}$ A is both practically and morally justified.

We are now in a position to state a rule that combines knowledge with morals:

NORM 10.4 Always choose the wisest of all possible courses of action.

We wind up this section with a sketch of a formal analysis of the concept of human action consonant with our ontology (Vols. 3 and 4). To begin with recall that all actions, whether human or not, are processes, i.e. changes in the states of some concrete thing or other. Call P the collection of all (lawful) processes, and A that of all possible human actions. Obviously, $A \subset P$. Human actions concatenate with one another as well as with natural processes. That is, if a and b are human or natural processes, their concatenations $a \cdot b$ and $b \cdot a$ are processes as well, i.e. members of P, though not necessarily of A. Symbolize inaction by '0', the identity element of P. By definition, 0 is the process such that, for any a in P, $a \cdot 0 = 0 \cdot a = a$. Because human actions combine with natural processes, the concatenation $\cdot$ is not closed in A but only in P. Hence this is the collection we must study. The simplest structure discernible in it is that of a semigroup with identity. Shorter: the totality of human actions is embedded in the structure $\mathscr{P} = \langle P, \cdot, 0 \rangle$, which is a noncommutative monoid. (It would be a mistake to think of $\mathscr{P}$ as a group because macroprocesses, which are the ones of interest in action theory, are irreversible. That is, it is not true that, for every action a, there is a counteraction $\bar{a}$ such that $\bar{a} \cdot a = 0$ or $a \cdot \bar{a} = 0$ capable of undoing the effects of a.)

However, the algebraic approach to human action is too coarse. For one thing it does not include time, whence the symbol '$a \cdot b$' is ambiguous, as it may denote either the joint or the successive performance of actions a and b. For another, the algebraic approach takes actions in themselves, for it makes no explicit reference to the two parties of any human action: the agent(s) and the patient(s). These shortcomings are overcome by adopting the systems theory formulated in Vol. 4, Ch. 1.

The basic concepts in our systems-theoretic treatment of human action are those of agent and patient, and their histories (or sequences of states) over a certain time interval. If we focus on the patient and neglect its reaction upon the agent, we only need consider the history $h_T(y)$ of the free patient y during a period T, and its forced history $h_T(y \mid x)$ when under the action of x. Each of these histories may be represented as an arc of a curve in the state space of the patient or, better, in that of the supersystem composed of agent and patient: See Figure 10.1. (Recall that a state space for things of some kind is the set of their really possible states, i.e. those compatible with the laws and constraints that characterize that species.)

Using the above concepts we make

DEFINITION 10.2 Let x denote an agent, and y an entity accessible to x.

(i) The *action* of x on y during the time interval T equals the (symmetric) difference Δ between the forced and the free trajectories (histories) of y in the state space of y (or of the supersystem composed of x and y):

$$A_T(x, y) =_{df} h_T(y \mid x) \, \Delta \, h_T(y);$$

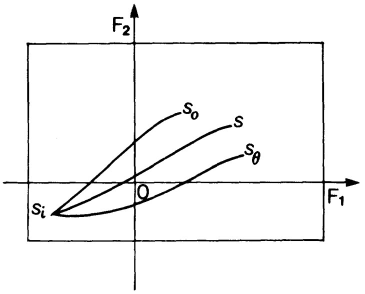


Fig. 10.1. Each of the three trajectories shown in the lawful state space S_L of the systems of the species in question represents a possible process undergone by a given specimen. (Actually the state space is a many-dimensional abstract space.) In one interpretation, each trajectory corresponds to a different strategy. In another, the upper curve, ending up at s_0, represents the history of the system while it is not being acted on by the agent. The lower trajectory, which finishes at s_θ, represents the theoretical or calculated trajectory, i.e. the one that the system would follow if it obeyed strictly the chosen strategy (perfect theory and no unforeseen external perturbations). The middle curve, ending up at s, represents the actual history of the system, which in general differs from both its free trajectory (no agent) and the calculated one. The efficiency of the action, or of the corresponding strategy, may be measured by some function of the difference between the actual s and the theoretical s_θ.

(ii) an action $A_T(x, y)$ is *effective* $=_{df}$ $A_T(x, y)$ contains the final or goal state s_f of y, i.e. $s_f \in A_T(x, y)$ — otherwise the action is *ineffective*;

(iii) an agent x is *inactive* with respect to all things of kind K during $T =_{df}$ For all y in K, $A_T(x, y) = \emptyset$.

1.3 *Goals, Means, and Plans*

Rational, responsible and minimally free people are not drifters but plan their activities and update their plans in the light of new events, fresh information, and changes in values and norms. Some of us plan our entire lives, usually with the help of relatives and friends, and attempt to live according to our life plans. There is nothing wrong with planning one's own life, and everything wrong with sticking to unfeasible or evil plans, or with allowing others to plan our own lives without our consent. All planning, to be efficient and morally right, must be participatory, i.e. it must involve all the people likely to be affected by its implementation. The prudential reason is simply that no individual is totally self-reliant and fully reliable: we are all dependent and fallible to some extent.

A plan or strategy may be analyzed as a function associating the ordered pair ⟨initial state, action⟩ with the goal state. More precisely, we propose

DEFINITION 10.3 Let s_i and s_f *denote the initial and final states respectively of a system under the action of a human agent. A feasible* (really possible) *plan* or *strategy* for performing such action is any function π such that

$$\langle s^i, m\rangle \overset{\pi}{\longmapsto} s_f,$$

where $m = A_T(x, y)$ is the action exerted by agent x on patient y during the time interval T in guiding y from s_i to s_f along a trajectory entirely contained within the lawful state space of y (or of the supersystem composed of x and y).

(For alternative conceptions of plans see Jantsch Ed. 1969 and Gale 1974.)

An important consideration in the choice of any plan is of course that of its efficiency. This typically technological concept may be characterized by

DEFINITION 10.4 A plan or strategy π is *efficient* $=_{df}$

(i) the means m of π is effective (on Definition 10.2 (ii)); and

(ii) the value of the goal s_f of π is much greater than that of its means m.

The choice of final state or goal is a technico-ethical problem. It is technical because it involves finding out whether the goal is attainable and affordable. And it is ethical as well because the change from the initial to the final state may have foreseeable harmful effects on someone, whether alive or to be born. On the other hand the design of efficient plans or strategies is a strictly technical problem best left in the hands of experts, in particular sociotechnologists — with the proviso to be noted anon.

Given a set of feasible plans or strategies, the problem of choosing among them the one(s) that will best agree with a given moral code is a strictly ethical problem. (Consequently the decision maker involved in projects likely to affect the well-being of many people assumes a heavy moral responsibility.) In our view the optimal plan or strategy is not just the most efficient one in view, in particular the one most likely to maximize the utilities of someone, but the one which satisfies

POSTULATE 10.1 The best plan or strategy for performing any human action A is the one most likely to satisfy, in the most efficient way, some of the basic needs or legitimate aspirations of all the people that may be affected by A, including its agent.

Though legitimate and useful, the means-end distinction is not absolute but situational, i.e. dependent upon circumstances. Sometimes the goal of an action is only a subgoal, i.e. a means for a subsequent action. And at other times the means turn into an end in itself, and the agent has to get hold of further means in order to attain it.

Example 1. We may work or fight for free speech and, having secured this right, we may exercise it as a means to defend or attack other rights — even free speech itself.

Example 2. The automobile, originally designed as a means of transportation, has become an object of worship, a status symbol, an obstacle to public transportation, and a health hazard.

Example 3. The struggle for power ought to be a means to a subgoal, namely the control of government for the sole purpose of helping people enjoy a satisfactory quality of life. Alas, all too often power becomes an end in itself.

Given that means and ends coevolve with society and change in value, it is unwise to stick dogmatically to any given set of means-goals pairs and, a fortiori, to any plans aiming at linking them. Instead, it is advisable to adopt

NORM 10.4 Goals, means and plans should be under constant technical and ethical scrutiny, which involves checking whether the goals are still worth pursuing, and the plans continue to be efficient, and altering either if they are not.

1.4 *Work*

All animals are active, but only humans earn their livelihood with the help of artifacts, whether primitive and material like scrapers and pebble axes, or sophisticated and symbolic like language and argument. Such activity is called 'work'. Work is distinguished from other kinds of activity in that it involves tools (material or symbolic) and is productive. Cooking a meal, mending a garment, designing a work schedule, laying bricks, and writing a paper for publication are instances of work. On the other hand eating a meal, going for a walk, reading a novel, studying for exams, and having a chat, are not.

All of the higher animals enjoy activity not only for the possible gains but for its own sake. For example, a hungry rat may prefer a new maze to going straight where it knows the food is; and a college student may prefer doing a hard and ill-paid job to being well paid for lying idly in bed all day long — as Hebb and others found long ago. As a result of his experiments in forced idleness, Hebb (1953) concluded that hedonism is psychologically false and proposed that "Man has an inherent need of work; that he is so made as persistently to seek a certain degree of discomfort, and persistently invite the risk of pain for the sake of that risk; and that a need for mental effort, including philosophic and religious thought, is built right into his nature from the first, as much as swimming is built into a fish". (See also Veblen 1899 for what he called "the instinct of workmanship".)

Why then the biblical belief that work is a curse? Why do so many economists, behaviorist psychologists, and utilitarian philosophers hold that man is basically a pleasure-seeker who will invest most of his energy in procuring pleasure, that he will learn only if he expects an external reward, and will avoid pain and effort at all costs? Why the persistence of the belief that only harsh criminal laws and a stern moral code, in conjunction with merciless competition in the labor market, will induce people to earn their livelihood by honest toil? The answer may lie in the working conditions of the slaves and serfs for which such laws and morals were designed. Heavy, joyless and often hazardous toil for the benefit of others is indeed a curse. So long as this is the case the worker feels alienated — as Hegel, Feuerbach and Marx noted — and the master feels contempt for labor. (Work did not occur in Aristotle's or Saint Paul's lists of virtues.)

Marx thought that alienation results from private property. Experience under socialism, and experiment under capitalism, have shown otherwise. Human engineers and social psychologists have found that alienation is caused by extreme division of labor joined with lack of participation. The former results in monotony — hence in boredom and fatigue — and the latter in the absence of responsibility or commitment. These hypotheses, suggested by observation, have been confirmed by experiments in job redesign, involving job rotation and a measure of self-government, at least with regard to the design of work schedules and norms. (See Buber Agassi 1986.)

What is usually called a *work ethic* is actually an attitude or value judgment concerning work. On the whole, both the leisure class and the lumpenproletariat despise work; blue collar workers hold manual work (though not necessarily workmanship) in high esteem but tend to look down on the work of people in other groups, particularly intellectuals; and scientists, engineers and artists tend to go by the so-called Protestant work ethic — which might as well be called Chinese or Japanese, Northern Italian or Northern French. (See Buber Agassi 1980.)

However, the work value system ("ethic"), and particularly the attitude towards workmanship, hence quality control, depends on working conditions as well as on occupation. Thus self-employed people and small work teams, who design their own work schedules and norms, usually find satisfaction in their jobs and therefore work hard and well. They do not work because they feel morally obliged to, or just because of the pay, but mainly because they like to do what they do: work is their main hobby. On the other hand people working on routine tasks, without chances of self-realization and growth, and with no say in the organization of their own activities, tend to be dissatisfied and consequently to work below the norm. (See Scitovsky 1976.) Hence the need periodically to redesign and reallocate jobs. This, and safety, are two areas where the interests of workers and management should coincide.

We summarize and complete the preceding by adopting

POSTULATE 10.2 (i) All and only work useful to self or society is valuable. (ii) Useful work is the most valuable of all human activities. (iii) Of all useful work, the one devoted to the public good is the most valuable. (iv) Work is both a basic human right and a basic human duty. (v) Every worker has the right and the duty to participate in the organization of his workplace.

By clause (i), work in the arms industry, the drug racket, or the pop culture industry is not valuable. By (ii) work should be exalted as a virtue. By (iii) government workers and volunteer workers, when they really act in the public interest, perform highly valuable work. By (iv) the employment rate (of women as well as men) is an index of social advancement. By (v) democracy in the workplace is a must. (See e.g. Vanek 1975, Zwerdlin 1980, Marković 1982.)

The recognition that work is a basic human right and duty entails that labor-saving devices should only be welcome when they do not contribute to unemployment. In particular, automation should be adjusted to our basic needs and legitimate aspirations, not the other way round. Robots should only be employed to eliminate hazardous or extremely boring jobs, never to create unemployment. (See K. Vonnegut Jr's *Player Piano* (1952).)

We compress the preceding into

Norm 10.5 (i) Everybody should work at what he does best. (ii) Nobody should be forced to take on a hazardous, exhausting, humiliating, or useless job. (iii) Working conditions should be constantly monitored and periodically revised so that they be technically, medically and psychologically satisfactory. (iv) Technology should only serve human welfare. (v) The first duty of industry is to serve the public.

We close this section noting that action theorists, with the sole exception of the Polish school of praxiology, have systematically ignored work. Why may this be so: general lack of academic interest in important problems, or aristocratic contempt for praxis?

1.5 *Summary*

Human action can be construed as the alteration, by the human agent, of the history of the system that is being acted upon, or patient. An action is effective if it leads the patient to the desired final state. And it is efficient if it does so in a manner that the resulting benefit by far outweighs the cost of the action.

Technologists, in particular sociotechnologists, study the way of optimizing the efficiency of deliberate human action. The moral philosopher is interested in putting such efficiency in the service of welfare. To this end he may propose certain moral norms, such as the injunction to refrain from unnecessarily destroying anything useful.

The wisest course of action is the one that is both the most efficient and morally the best. Preparation for such action involves the judicious choice of the goal and the design of a plan or strategy likely to lead to the goal in the most efficient way compatible with the moral constraints.

Work, particularly in industry, is a prime example of rational and planned action. Far from being a curse, work is, or can be made into, a source of pleasure; when not, it must be redesigned or reallocated. In any event, work is both a basic human right and a basic human duty. But not all work is morally justifiable; for example, work on the design or manufacture of offensive weapons, or of useless trinkets, is not morally justifiable. It we wish to preserve our industrial civilization we must subject industry to certain constraints aimed at preventing industry from harming our health and deteriorating the environment. Industry is an admirable servant but a redoubtable master. Let it be a servant of humankind.

2. COLLECTIVE ACTION

2.1 *Private and Public Interests*

Human actions are performed to further the interests, real or apparent, of individuals, groups, or entire societies. (Disinterested actions are no exception: they further the interest of someone other than the agent.) These interests are biological, economic, cultural, or political. Although these concepts are familiar, they are important enough to warrant explicit definition. We propose

DEFINITION 10.5 A decision or action A is of

(i) *private interest* $=_{df}$ A is intended to affect the well-being of an individual;

(ii) *special group interest* $=_{df}$ A is intended to affect the well-being of the members of a social group other than a family;

(iii) *public interest* $=_{df}$ A is intended to affect the well-being of most members of a society.

Obviously, not every decision or action *of* interest to X is *in* the best interest of X. We submit that we should do our best to engage only in actions that are morally justified as well as expected to be practically efficient. But first another definition, which specializes the rather noncommittal Definition 10.1 (iv) in Sect. 1.2:

DEFINITION 10.6 A decision or action A is *morally justified* $=_{df}$ A is intended to contribute to the well-being of the people it is likely to affect, without preventing anyone from meeting his basic needs.

And now a principle:

NORM 10.6 (i) We should only take morally justified decisions or actions. (ii) A decision or action A taken to further the public interest prevails over both special and private interests if and only if A is morally justified. (iii) A decision or action A taken to further a special interest overrides private interests if and only if A is morally justified.

Obvious examples of morally justified actions are raising taxes to finance public services, confiscating guns, jailing and reeducating criminals, manufacturing useful products, and exposing quacks. Of morally unjustified actions: raising taxes to finance an aggressive foreign policy, jailing political opponents engaged in peaceful activities, censoring scientific research, and circulating malicious gossip. Note that, whereas some of these actions are individual, others are collective.

Individuals can influence the state of society by acting singly or jointly. They do the former when they buy a consumer item or when they vote, the latter when they work in a firm or a government, or when they participate in a collective action such as a strike or a political rally. We must not think of collective action in a holistic fashion, i.e. as the action of a body of people as if it had a purpose of its own, but in accordance with

DEFINITION 10.7 An action is *collective* $=_{df}$ There is a social group all the members of which engage in individual but coordinated action with the same purpose, though not necessarily with the same motivation.

More precisely, using Definition 10.2 we may construe the action of a social group G_1 on the social group G_2 as the union of the individual actions of the members of G_1 upon those of G_2 during a given period T, i.e.

$$A_T(G_1, G_2) = \bigcup_{\substack{x \in G_1 \\ y \in G_2}} A_T(x, y).$$

Note that, although every social action is the "sum" of individual actions, the total outcome may be a radical alteration of the social structure of the two groups involved: think of social revolutions, whether violent or quiet.

The solo recital and the orchestra performance exemplify individual and collective action respectively. Collective action is characterized by cooperation and, if it is to be effective, takes planning and organization.

Even individual action may require some planning and the cooperation of others. For example, a soloist cannot do without the help of a manager, some technicians, and a number of anonymous and unwitting helpers, such as letter carriers and newspaper vendors — unless of course he only plays for a small circle of friends.

Cooperation is all the more necessary when the goal is to solve some social problem. The reason is obvious: Social problems are many-sided and affect many people, whence their solution involves large human and material resources. Loners wield no power. Regrettably much contemporary social science is concerned with isolated individuals and overlooks collective action — as a result of which it sheds hardly any light on social problems. A good example of a well-meaning but ineffectual academic exercise is social choice theory, which purports to map individual preferences into social preferences. (e.g. Arrow 1951, Sen 1970, Bonner 1986). Though exact, this theory is shallow, unrealistic, and ridden with paradox. It is superficial for focusing on subjective preferences, which need not reveal objective needs; and it is unrealistic for assuming that individual choice is always free. The first flaw results from the positivistic rule that the scientist should stick to observable facts, in this case choice behavior. But in real life choice, unless it is impulsive rather than rational, is only the last link of a chain that the scientist ought to investigate in its entirety: Need or desire-Consciousness of need or desire-Survey of available or procurable means-Means evaluation-Risk evaluation-Choice. As for the assumption that everyone is free to choose anything, it overlooks constraints and commitments of all kinds (in particular moral duties), and it ignores the fact that we often abstain from choosing because we do not like any of the options offered us. Because of these basic flaws, we need not dwell on the center-piece of the theory, which is Arrow's paradoxical "general impossibility theorem", according to which social choice cannot be both rational and democratic.

To reduce politics to voting, and political problems to devising workable electoral schemes, is just as simplistic as to reduce the economy to the choice of breakfast cereal, and economic problems to its pricing or marketing. Social problems pose the conceptual problem of designing organizations or even entire social orders capable of coping with them, and the practical problem of finding or producing the means to set up such organizations or effect such social restructurings. The former or theoretical problem will be tackled in the following chapter. Let us now deal summarily with some of the conceptual issues raised by the latter or practical problem.

People who wrestle with social problems join or set up organizations and, by participating in them, engage in collective actions. Such organizations may be governmental, private, or mixed; they may further the interests of groups or of entire societies; and they may be transient like committees, or permanent like unions or clubs. The design and assembling of an organization presupposes that its possible value has been assessed beforehand; and the continued existence of the organization calls for its periodic reevaluation. Such evaluations may be helped by

DEFINITION 10.8 Let S be a society, G a special group of S, and O an organization in S. Then

(i) O is *valuable to* $G =_{df}$ on balance O is beneficial to all or most of the members of G (i.e. the benefits accruing to nearly every member of G from belonging to G outweigh its costs);

(ii) *O is valuable to* $S =_{df}$ (a) O is valuable to G and (b) O does not hinder any member of S from meeting her basic needs and legitimate aspirations.

For example, whereas a public utility and the traffic police are, if honest and efficient, valuable to all the society they serve, a private army or club, as well as a church, are only valuable to its members. This is why tax payers have the duty to support the former but not the latter. Special interest organizations should be viewed with favor only if they do not act against the public interest. Sometimes they undermine public interest organizations, to the point of causing the decline of entire societies (Olson 1982). This suggests

POSTULATE 10.3 An organization is the more valuable the larger the number of people whose well-being it helps maintain or improve without thereby hindering anyone from meeting his basic needs.

The proviso is needed to disqualify organizations that thrive on the exploitation or even extermination of national minorities or foreign peoples.

One of the problems posed by the design or redesign of any organization is whether it should be centralized or decentralized, authoritarian or democratic. Since these concepts are somewhat tricky we may as well define them:

DEFINITION 10.9 An organization O is *centralized* $=_{df}$ O contains a single component or subsystem, its *center* or *nucleus*, on which the activities of all the other components of O depend. Otherwise, i.e. if it contains either no such center or two or more centers, O is *decentralized.*

Examples of centralized systems: an atom and a firm managed by its owner.

DEFINITION 10.10 An organization O is

(i) *authoritarian* $=_{df}$ the relation between the center of O and the other members of O is that of one-sided subordination (no feedback);

(ii) *semidemocratic* $=_{df}$ the members of O other than its centers obey the latter but there are recognized mechanisms (such as suggestion boxes, quality control circles, and grievances committees) allowing them to react (information feedback);

(iii) *democratic* $=_{df}$ all the members of O participate, either directly or by delegation, in the activities of the centers of O.

Clearly, centralization and democracy are mutually independent. An organization can be centralized but democratic, and polycentric but authoritarian. Moreover there are degrees of decentralization; for instance, the links of a supermarket chain may or may not have local autonomy. Likewise there are degrees of democracy, from representative to participatory. See Figure 10.2.

The choice of organization is technically and morally important. Efficiency is best achieved by decentralization and participatory democracy — as long as they are combined with technical expertise. Decentralization enhances agility, participatory democracy facilitates involvement and cooperation. And both together stimulate personal initiative and responsibility. Moreover they alone can prevent exploitation and corruption.

However, decentralization and democracy can lead to stagnation or

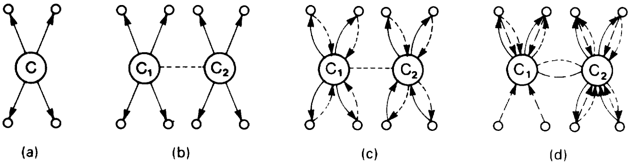


Fig. 10.2. Four types of organization. Central circle: management. Small circle: rank and file member. Full line: power relation. Dotted line: information link. Broken line: participation. (*a*) Centralized authoritarian. (*b*) Decentralized authoritarian. (*c*) Decentralized semidemocratic. (*d*) Decentralized democratic.

even chaos unless combined with technical expertise. This particular combination is so important that it deserves a name of its own:

DEFINITION 10.11 A social system S is a *technodemocracy* $=_{df}$ S is democratic and its policies, plans and decisions conform to up to date technology.

Technodemocracy should not be mistaken for technocracy, or the rule of technologists. The latter should be neither on top nor merely on tap, but should actively participate in the democratic process of designing and running an organization or a society.

In a technodemocracy people are reasonably well-informed about the business of their organization and they get involved actively in its decision-making process instead of limiting themselves to casting a vote, once in a while, for a professional politician or labor organizer. By becoming involved they do not feel alienated but like their work and are ready to help their coworkers. Technodemocracy is thus an effective means for implementing the basic norm *Enjoy life and help live.*

2.2 *Social Policies and Plans*

A policy is a set of principles to guide plans, decisions and actions of some kind. (Barbarism: 'philosophy'.) For instance, each university has its own admissions policy, and each government its own cultural policy. A *social policy* is a policy aiming at solving social problems, i.e. problems concerning an entire social group or even a society.

A good social policy is realistic and has a socially valuable goal. To be both realistic and socially valuable, a policy must be *technically sound* rather than amateurish, *global* rather than sectoral, *long-sighted* rather than myopic, *participatory* rather than authoritarian, and it should seek to *improve the well-being* of the greatest number instead of securing privileges for the few. We take this last condition for granted in view of the values and morals advocated earlier in this book. Let us justify the four preceding conditions.

A social policy is bound to be ineffectual or worse unless it is designed with the help of the best available knowledge: enthusiasm and dedication are no substitute for the scientific knowledge of social mechanisms. Moreover, given our scant knowledge of the latter, every important new policy should be tried out experimentally on a sample group before being applied to the whole. Only such a test can indicate whether the proposed policy does, in fact, effect what the policy-makers seek. Governing in the light of social experiment is slower but more responsible and less wasteful than governing by decree.

Social policies should be global, not sectoral, because social problems are many-sided, and this because every person is not only an animal but also a member of a number of social groups. For example, it is unrealistic to promote public education without at the same time improving the health and income of the more disadvantaged people, for a sick or hungry child or youngster is a poor learner. Likewise any purely economic policy of economic growth is unrealistic, because modern industry can employ only workers capable of understanding written instructions and of learning new skills. As well, modern industry requires an institutional framework and a government that allows it to conduct their business in an efficient manner.

The need for long-term policies becomes obvious when one realizes that most social problems have deep roots in the past, and that whatever is done to solve them is bound to take a long time and to affect future generations. For example, the eradication of certain plagues and of illiteracy may be accomplished in one decade, but that of poverty, power abuse and moral corruption may take several decades of hard and persistent work. There are no quick fixes for social problems.

Finally, the involvement of all the interested parties in the policy making process is necessary for practical as well as moral reasons. The former because only people who participate in that process feel responsible and accountable for the implementation of the policy, hence take their job to heart and devote it their effort and imagination. The moral reason is that authoritarianism violates human rights and divides society into moral agents — the rulers — and moral pawns — the ruled.

We compress the preceding into

POSTULATE 10.4 A good social policy is a global (or systemic) and long term policy aiming at increasing social welfare, and designed with the help of the best available relevant knowledge, as well as with the participation of the people likely to be affected by its implementation.

Now, a policy without plans to implement it is like a constitution without legal codes, that is, a rhetorical statement of intention. If one intends to have a policy implemented, he will insist that it be accompanied by realistic plans. A plan or strategy is of course a prescription for guiding the evolution of a system from its present state to a prescribed goal state. (Recall Definition 10.3.) Just as reflex and impulsive action are typically animal, so planned action, particularly when the plan is rational and long-sighted, is distinctively human. However, we still tend to be improvident in such vital areas as human reproduction and environmental management. This deficiency is in marked contrast with the detailed military plans for annihilitating all life on our planet.

The general concept of a plan or strategy was defined in Sect. 1.3. A *social plan* (or *program*) is a prescription for guiding (goading or forcing) the evolution of a large-scale social system according to a certain social policy. A social plan may be local, national, regional, or world-wide, according to the size of the corresponding social problem and the vision of the decision makers. In line with Postulate 10.4 we propose

POSTULATE 10.5 A good social plan is one designed to implement, in a participatory and technically adequate manner, some aspect(s) of a good social policy.

Good social planning is the exception: most social plans are sectoral, myopic, authoritarian, and either too timid or too radical. In particular, most national development plans have been designed by myopic economists who equate progress with increased aggregate economic activity, and have consequently neglected the biological, cultural and political components of development. In the best of cases such one-sided plans have brought about ephemeral bouts of industrial growth — but at the expense of the environment, public health, education, culture, and civil liberties. Since society is a system composed of a number of mutually dependent subsystems, neither of which is the prime motor, it cannot progress as a whole unless all of its components advance synergically at the same time. A bad social plan can be worse than no plan at all.

Hayek (1944) and his fellow individualists oppose all kinds of social planning, holding that it is “the road to serfdom”. However, there is nothing wrong with planning *per se*, which is “the orchestration of life”, as Neurath (1945) wrote in reply to Hayek. It all depends on the goal and the means used to attain it. A plan to plunder or conquer a foreign nation, and the more so the whole world, is evil. But a technically adequate plan to improve social welfare by democratic means, or to put an end to the arms race, or to environmental degradation, is good, nay necessary for the survival of humankind.

It is not just that plans are necessary to orient individual and collective action in the contemporary world. Good planning is the road to freedom from want as well as to freedom to become what one wants without trampling on the rights of others. Slaves need not plan: their masters plan their entire lives for them. Nor do derelicts: they live, or rather die slowly, from day to day. But everyone else and every social organization needs some plan of action. Improvidence is imprudent for leaving one's life, or that of one's organization, at the mercy of others. And it is morally objectionable as well for involving the neglect of duties and the relinquishing of right.

Now, global or systemic plans are necessarily central: they must coordinate local plans in order to secure efficiency and a fair distribution of the available resources. And such centrality is bound to strengthen the power of the ruling élite at the expense of local autonomy and individual freedom. Hence social planning would be bad after all. However, this conclusion depends on the premise that the government is not democratic, whence its social plans are not participatory. We have insisted before that a good social plan is participatory. Such a plan will not seek to rule in minute detail the life of every individual, but to help him find his own way without exploiting or oppressing others. Moreover, a participatory plan will not be a final blueprint but rather a sequence of blueprints designed to cope with changing circumstances and new information: it will be fluid rather than rigid. (See Vol. 7. Ch. 5, Sect. 1.1 and Ackoff 1974.) In sum, the question is not whether we need plans, but only what kind of plans will best suit the individual and public interests. The general answer to this question is provided by Postulates 10.4 and 10.5.

Let us finally note that, just as any realistic budget will include a slot for unforeseen expenditures, so any realistic social plan will make room for unforeseen events. These are unavoidable because of the very nature of large systems and of the imperfection of our knowledge of them. In particular, one must be prepared to cope with natural catastrophes of known kinds, such as droughts, and one must make room for the unintended consequences of our actions, particularly when not well-coordinated. As well, we must allow for changes in values, for our insufficient knowledge of the present state of the world, and for the dearth of dynamical models of society which enable us to make reliable social forecasts beyond a few months. Still, the answer to the unforeseen and to ignorance is not drifting but research and participatory planning.

In short, policies and plans are indispensable because they are, at least in part, self-fulfilling forecasts. The future belongs to those who design and build it rather than to those who wait for its arrival.

2.3 *Cooperation and Competition*

Cooperation, of which group action is but an example, is pervasive in all animal societies. Without it there would be no social groups, and without these the individual would be lost, for nobody is totally self-sufficient. Cooperation is then an essential survival mechanism. Therefore individualism, and particularly egoism, go against the grain of biology and sociology.

As a matter of fact cooperation occurs on all levels of reality. Examples of cooperative processes are condensation, molecular synthesis, cell self-assembly, and the formation of biological colonies and communities, as well as the emergence of human groups. But competition too is pervasive in nature and society. Examples: the competition of two molecules for a third in the course of two concurrent chemical reactions, of two cells or multicellular organisms for a nutrient, of two males for a female, of two firms for a market, of two countries for a third. Given the pervasiveness of both processes we may as well define the corresponding concepts in general terms:

DEFINITION 10.12 If a and b are two concrete things,

(i) *a helps* (*obstructs*) *b* in the respect(s) $R =_{df}$ some of the functions (activities, processes) of a facilitate (hinder) the function(s) of kind R of b, and none blocks them in the same respect;

(ii) *a cooperates* (*competes*) with $b =_{df}$ a helps (obstructs) b in some respect(s) and conversely.

Note the following points. Firstly, neither cooperation nor competition requires intention or consciousness: either of them may occur among inanimate systems, organisms lacking an advanced nervous systems, or social systems. Secondly, cooperation in a given respect is compatible with competition in another. Thirdly, we do not distinguish competition from conflict.

We cooperate with other people for a number of reasons: out of sympathy or empathy; for the pleasure of feeling that we belong in a group; out of self-interest, i.e. for reckoning that such behavior may result eventually in some advantage for ourselves; from a sense of duty; or from fear of social sanction. The multiplicity of solidarity sources refutes the postulate of individualistic ethics and individualistic social science, that we only act from self-interest. It also helps dissolve the pseudoparadoxes encountered by social scientists blinded by individualism: Why most people do some favors without expecting anything in return; why most of us are not criminals; why we normally prefer reciprocating to defecting; and even why we often vote knowing that a single vote is unlikely to make any difference. All prosocial behavior, in particular cooperation, is bound to look paradoxical to the individualist, for he believes that people are only moved by self-interest. This is the case with all the game theorists who equate rationality and selfishness. However, a few students have used the very same theory to prove that cooperation pays more than defection, at least when the probability of a next encounter is not negligible. (See e.g. Rescher 1975, Axelrod 1984, Alexander 1987.) Since the same theory is being used to derive certain theorems and their negations, there must be something basically wrong with it. We claim that the trouble with decision theory and its applications is that it employs subjective values and probabilities: Recall Ch. 3, Sect. 3.3. Let us go back to real life.

The family, the voluntary or non-profit organization, and the cooperative firm, are good examples of the benefits of cooperation. Let us take a quick glance at cooperatives, of which there are several thousands throughout the five continents. One of them, the Mondragón complex in the Basque Country, has recently attracted the attention of a number of social scientists (e.g. Thomas & Logan 1982). Its remarkable sucess has been attributed to the following factors: (*a*) Mondragón is a system with more than 100 components, mostly factories, distributed among several sectors of the economy; (*b*) it has its own bank, so it does not depend on a hostile financial system; (*c*) it has its own R&D centers — which keep it technologically up to date — and its own B&A schools, which train and recycle its managers; (*d*) it is decentralized, hence flexible; (*e*) it is self-managed, hence it provides work satisfaction and its concomitants; (*f*) its members have a pecuniary stake in it, and (*g*) its founders and leaders have aimed at service rather than just profit. There are similar success stories elsewhere, which suggest that economic democracy and self-management are not utopian. We shall return to self-management in Sect. 2.5.

Cooperation is one side of the social coin; the other is competition, which is just as pervasive. Dialectical philosophies notwithstanding, competition is not confined to pairs of opposites. Rather, whenever the available resources are limited, similarity of needs or interests leads to conflict. In other words similars, rather than opposites, are likely to compete for any given natural or social niche. Competition is not just unavoidable: it is also necessary to promote competence, quality and productivity, because it stimulates initiative, creativity and industriousness. Without competition, mediocrity and complacency are bound to set in, and consequently stagnation or decline are bound to occur.

Now, all competition hurts, and unbridled competition destroys individuals and entire social groups. Therefore it must be kept under control. The moral reason for controlling competition is obvious: we must help others to live. The prudential reason is no less evident: the concentration of power is only good for those who wield it, and it elicits revolt, which may be disastrous for everyone.

Conflicts can be resolved by force or by peaceful means. The latter are preferable for being less wasteful, avoiding revenge, and observing the supreme moral principle *Enjoy life and help live*. The main peaceful method of conflict resolution is negotiation or bargaining, which combines confrontation with cooperation. Bargaining is of course a process wherein every proposal or demand is met with a counter-proposal or concession, until the last demand is satisfied — or not, in which case the conflict is handed over to an arbiter, or left unresolved.

Bargaining is of interest to the moral philosopher on several counts. Firstly, is it ever justified to refuse to negotiate? In our view never, particularly when human lives are at stake. Secondly, what is negotiable and what is not? In our view only life is not negotiable; everything else, including one's own well-being, is negotiable. Thirdly, what is a fair deal? Answer: A deal is fair if and only if one party's losses are compensated for by its gains. When the mutual concessions can be quantitated, a fair deal is one wherein the value of each party's concessions is the same as that of the other. Unless such balance obtains, the deal is unfair, i.e. one of the parties takes advantage of the other. The best deal is of course that where the two parties stand to gain while no third party is harmed.

In real life the outcome of a bargaining process is usually determined by the relative bargaining power or leverage of the parties concerned. For example, workers have little bargaining power when they are not unionized, or when the unemployment rate is high, or when they compete with machines; and, unless they join forces (as e.g. in the OPEC), the less developed nations have hardly any bargaining power *vis-à-vis* the industrialized ones. Because such inequalities give rise to unfair deals, the contractarians (e.g. Gauthier 1986) are mistaken in holding that contracts exude morals: very often they just legalize predation.

However, in the long run immoral deals do not pay. In fact, an unfair deal is likely to cause the resentment of the weaker party, which may eventually repeal the contract and seek redress by stealth or by force.

(For example, the Nazis used the unfair Treaty of Versailles as a bait to persuade the German people to go to war.) Because unfair deals are likely to be short-lived and the seed of discontent, a prudent negotiator will only strike fair deals. Prudence and morality go hand in hand.

2.4 *Competitive Cooperation*

Cooperation is not good in itself: its value depends upon its goal. Organized crime, lynch mobs and aggressive armies furnish examples of antisocial cooperative behavior. (See Olson 1982.) And competition is not always bad. For example, peaceful competition over excellence in learning, productivity, altruism, or sports, is good.

Nor is either conflict or synergy the driving force of every system in every circumstance. Sometimes competition is the engine, at other times cooperation, and at still other times, a combination of the two. (See Bunge 1981.) Such a combination is all too often overlooked. Yet it is well known that, in order for a system to compete successfully, it must be internally synergic to begin with: internal division weakens any team. Still, such synergy may and need only be partial. Indeed, in order for the members of a social system to enjoy belonging to it, they must be able to pursue their own interests, which are likely to conflict in some regards with those of some fellow members. (Think of a family or of a chess club.) Hence cooperation in some respects may coexist with competition in others. This is why Definition 10.2 relativizes cooperation and competition.

It follows that the dialectical or conflict-centered philosophies of society, like Marx's, are just as one-sided as the synergistic or solidarity-centered ones like Durkheim's. Since each view illuminates only one side of the social coin, both are inadequate for understanding it and, a fortiori, for altering or controlling it in a rational and effective way. These views are not only one-sided but also self-contradictory. Indeed, orthodox Marxists are supposed to favor a classless society and a peaceful international order, but according to dialectics both would be stagnant for lack of conflict. On the other hand some of those who preach class solidarity also uphold class stratification, and occasionally aggressive nationalism as well.

Since in real life we engage in cooperation as well as in competition, and since each of them has its positive and its negative aspects, we should explore how best to combine these two modes of social behavior so as to prevent antisocial action. The most general exploration of this kind is mathematical, for it overlooks the peculiar properties and a fortiori the laws of the entities concerned. We proceed to presenting the bare bones of extremely simple (hence somewhat unrealistic) cases of pure cooperation, pure competition, and competitive cooperation.

Consider two entities, to be called 1 and 2, characterized by certain time-dependent functions A and B respectively, which may be thought of as the amounts of things consumed or produced by the entities concerned. We shall postulate certain simple relations between A and B and their respective rates of change $\dot{A}$ and $\dot{B}$, and shall derive the main consequences in each of the three cases.

(i) *Pure cooperation*

$$\dot{A} = aAB, \quad \dot{B} = bAB, \quad \text{with} \quad a, b > 0.$$

Upon dividing the second rate equation by the first, and integrating, we obtain the solution of the system in the state space formed by A and B:

$$B(t) = (b/a)A(t) + c, \quad \text{with} \quad c = B(0) - (b/a)A(0),$$

where $A(0)$ and $B(0)$ are the initial endowments of the parties in question. In short, the prosperity of each component is proportional to that of the other. Examples: a cohesive soccer team and a cooperative.

(ii) *Pure competition*

$$\dot{A} = a(A - B), \quad \dot{B} = b(B - A), \quad \text{with} \quad a, b > 0.$$

That is, the growth (or decline) of each component is the faster, the greater their difference. The solution of the system of equations in the state space $A \times B$ is

$$B(t) = -(b/a)A(t) + c, \quad \text{where} \quad c = B(0) + (b/a)A(0).$$

Each of the components can only gain at the expense of the other. Furthermore, which component will prevail is determined by the initial endowments. Examples: competition between a small and a big firm, and between an industrialized and an underdeveloped nation.

(iii) *Competitive cooperation*

$$\dot{A} = a(\alpha - \mathrm{A})B, \quad \dot{B} = b(\beta - B)A, \quad \text{with}$$
$$a, b > 0, \alpha > A(0), \beta > B(0).$$

Component 1 helps 2 in proportion to the latter's need or deficit, i.e. in such a way that the imbalance $\beta - B$ will be offset. In turn, component 2 contributes to 1 in proportion to the imbalance $\alpha - A$, though not neccessarily with the same intensity (i.e. a and b may be different). This mutual help continues as long as there is a deficit or negative balance. It ceases on attaining satiety, i.e. when $A(t) = \alpha$ and $B(t) = \beta$, and it becomes negative as soon as there is a superavit, i.e. when $A(t) > \alpha$ and $B(t) > \beta$. That is, as soon as one of the parties starts getting more than it needs, the other starts opposing it, thus inhibiting its growth. If both parties have the same requirements (i.e. $\alpha = \beta$) and each of them helps its partner as hard as the other (i.e. $a = b$), both do equally well regardless of their initial endowments. (For a detailed mathematical study see Bunge 1976.)

This mathematical model suggests designing social systems in which every member helps his fellow member in proportion to his own ability as well as to the other's needs, while controlling the latter's growth and being in turn controlled by it. In this way domination, scarcity and waste would be avoided by combining the best of cooperation with the best of competition.

In principle one might think of the dual combination, of internal competition with external cooperation. Such a combination is feasible as long as internal competition does not lead to the collapse of the system.

2.5 *Management*

The management of an organization is the regulation, in particular the coordination, of the social behavior of its members in view of the efficient performance of the tasks of the organization. (See Vol. 7, Ch. 5, Sect. 4.1.) Management may be authoritarian or democratic, traditionalist or scientific, but it must be there if the organization is to function as a well-integrated system. Even the management of natural resources is ultimately the management of people: the regulation of the way they exploit, renew or enjoy the resources. Therefore Proudhon's famous saying, that after the next social revolution politics would become "the administration of things not of people", was mistaken. So was the belief of Lenin and Mao that management could be improvised or even replaced with revolutionary enthusiasm.

Because management controls behavior, it allocates duties and may restrict rights. Consequently management poses problems that are at the same time practical and moral. Two of these problems are: Who should manage, and Which is the best management style? Technical expertise, e.g. the use of operations research, is certainly part of the answer. But it is not the whole answer, because the most efficient management is that which favors the active participation of all the interested parties, for people function best when they become involved and feel responsible. Moreover, participatory management is not only efficient: it also satisfies the principle of qualified equality (Ch. 6, Sect. 2.2). In short, technodemocracy, or the combination of expertise with democracy (Definition 10.11), is the best answer to the technico-ethical problems of management.

Participatory management comes in various strengths. The weakest form occurs in privately owned enterprises where each work group determines its own work schedules and job rotation schemes, and studies proposals for improving product (or service) quality and job quality. (See Herbst 1974 and other publications of the Tavistock Institute in London, Buber-Agassi 1986, and Halal 1986.) The strongest form of participative management occurs in cooperatives owned and run by the workers themselves, of which there are more than 600,000 in the world. (See e.g. Vanek Ed. 1975, Miller 1977, Selucky 1979, Turkish Comisso 1979, Clayre Ed. 1980, Schweikert 1980, Zwerdlin 1980, Marković 1982, Dahl 1985.)

All organizations face the size problem. The classical view, defended by orthodox socialists as well as by procapitalists, is summarized in the slogan “Big is beautiful”. The anarchist view is that “Small is beautiful”. Actually neither giantism nor dwarfism is good or bad in itself: The optimal size of an organization depends on its specific tasks. Some social systems, such as day care centers, corner grocery stores, and watch assembly plants, are best kept small; others, such as steel mills, large city water works, and international organizations, are necessarily large. However, if we seek to optimize job satisfaction and participation as well as efficiency, we had better seek the division of every large enterprise into manageable subsystems, and the integration of all small enterprises into well-coordinated federations. Every unit should be big enough to function efficiently, and small enough to manage itself.

The problem of the size of a cooperative has a moral aspect, namely this. If the number of cooperants is too small, they will exploit themselves; and if it is too large, most of them will be left out of the decision-making process, and some of them will be free-riders. But even after having achieved the optimal size, this size may have to be altered to make room for technological changes. Hence planning should be flexible and, in particular, it should allow for splittings and mergers. In any event, technical considerations, e.g. with regard to efficiency, go hand in hand with moral ones, such as the avoidance of exploitation, in particular parasitism, and the achievement of full employment.

What holds for the size of economic units, such as cooperatives, also holds, *mutatis mutandis*, for cultural and political systems. In particular, it holds for the organizations in charge of the management of public goods, i.e. governments. The size of a government should depend exclusively on the social services it performs, and on the extent to which the civil society it manages is democratic, i.e. organized into self-governing units.

A government that facilitates the self-government of all organizations, while at the same time protecting the basic rights of all individuals, can be trim, hence unoppressive and inexpensive. Strong governments are necessarily big, hence oppressive in some way or other, and expensive. This is not to say that we should aim for a minimal state the sole function of which is to protect privilege. Instead, we should aim for a minimal state which facilitates the pursuit of individual well-being and the survival of humankind.

The survival of humankind is jeopardized by a number of global problems, for which no national government has the right or the power to impose global solutions. Only a world government could possibly tackle such world-wide problems. The world government should confine itself to managing the natural and cultural heritage of humankind. It would be impractical and immoral for it to attempt to control every single social group on earth: it would have to respect the authority of every member government over its own jurisdiction. But at the same time national sovereignty would have to be limited so as not to put the survival of humankind at risk. In international relations, like in interpersonal relations, the norm should be: Your freedom ends where mine begins, and the freedom of all can only be attained through equity, mutual help, and respect for international law. The United States of the World may now seem a naive utopia, but it will have to be established gradually if the human species is to survive. (See Albert 1985 on the roles of utopias.) We shall return to this matter in Ch. 11, Sect. 4.

The general management principles sketched above are summarized in

NORM 10.5 (i) All management should be technodemocratic, i.e. it should combine technical expertise with self-government. (ii) All of the organizations of a given kind, formed by individuals, should join to form federations on various levels (e.g. local, regional, national, and international). (iii) The preceding principles should be applied, in particular, to governments. (iv) The task of a world government would be to work on global issues, i.e. world-wide problems that cannot be solved by any national government.

2.6 *Summary*

Collective or group action is coordinated individual action. It is co-operative and, when efficient, it involves planning and organizing: spontaneity is a guarantee of chaos and failure. Organizations can be centralized or decentralized, authoritarian or democratic, and they can be run traditionally or scientifically. The combination of decentralization and democracy with technical expertise may be called 'techno-democracy'. Technodemocracy is practically desirable because it benefits from sociotechnology and, by involving most of the members of the organization, it gives them a stake in it. And it is morally desirable because it makes demands on individual responsibility and gives everyone a chance to protect his own interests and to help others do their job in a way convenient to all.

Social issues are tackled in the light of social policies. A good social policy is scientifically sound, technically feasible, and it aims at promoting the public good through the protection of legitimate individual interests. It is global or systemic not sectoral, long-sighted not myopic, and participatory not authoritarian. A social plan or program is a prescription for guiding the evolution of a social system in line with some social policy. In itself, planning is morally neutral: its value depends on the values of its goals and means. Some plans promote welfare and liberty, others misery and oppression. Good social planning is a means to attain freedom from want as well as the freedom to become what one wants to be without causing harm to others.

Social life has two sides: cooperation and competition. Without the former there would be no social groups, and without the latter progress, if at all, would be sluggish. To enjoy life we must combine cooperation with competition in such a way that the former does not lead to stagnation or decline, and the latter to destruction. Cooperation within a group or between groups is necessary but it may be limited in some respect, while there is competition in others. We combine cooperation with competition whenever we engage in a bargaining process to resolve a conflict of interests. Another way of practicing competitive cooperation is to stimulate cooperation as long as neither of the parties concerned has achieved its legitimate goals, and to promote their competition thereafter.

The management of every organization faces, among others, the problem of optimal organization size, which is ethical as well as practical. The size of an organization should be determined by its tasks and internal structure. Neither smallness nor bigness is good or bad in itself. However, in order to facilitate self- management all large organizations should be split into smaller though mutually linked units. The slogan might well be the devise of a Montreal drugstore: "Big enough to serve you well, small enough to get to know you". However, such decentralization should be accompanied by regional or even world-wide federation.

The global issues now facing humankind can only be faced by world wide organizations. Eventually a world government will have to be established to manage the common heritage of humankind and to prevent its extinction. It will be one world or none.

CHAPTER 11

SOCIAL PHILOSOPHY

We are judged by what we do, but we act according to our beliefs, values and norms — or those of our masters. When we render our social beliefs, values and norms explicit, and organize them into a coherent system, we obtain a social philosophy or ideology. A social philosophy will be sound or unsound, depending on whether or not it is based on science and technology, as well as on a science-oriented ontology, epistemology, axiology, and ethics. And it will be good or bad, depending on whether it enhances or jeopardizes everyone's chance to enjoy life and to help others live. Hence we should not underestimate the power of social philosophy: it can save or destroy us. True, most policy and decision makers do not read philosophers unless they are bad or dead. Still, every social policy has an underlying social philosophy. Hence it behooves social philosophers to watch the political process and to attempt to influence it in the right direction, instead of playing game-theoretic games.

Every social philosophy deals with social issues. And every social issue has a moral aspect, i.e. it involves a moral problem in addition to a practical one, for it concerns rights and duties rooted in needs and wants. At first sight such problems as devising an efficient method of tax collection, mail distribution, waste disposal, ballot counting, or the teaching of mathematics, are purely technical problems. On second thought every one of these problems raises ethical questions, such as those of loopholes in tax laws, unemployment caused by automatic mail sorting, environmental deterioration caused by most current waste disposal procedures, fraud that may be committed in ballot counting, and discrimination against the economically and culturally deprived in the learning of mathematics. In short, all social issues pose ethical problems, whence they cannot be resolved without appealing to some (explicit or tacit) ethical principles in addition to the findings of social science and sociotechnology.

Because of their magnitude, social issues call for collective action (Ch. 10). In turn, group action raises ethical problems, such as whether we are morally justified in being either conformists or revolutionaries in view of the facts that conformism condones much misery, and revolutions may be new sources of unhappiness. In line with our ethical views, we favor permanent nonconformism and permanent social reform. We favor nonconformism because no society can be perfect, and no individual can remain morally clean if he does not rebel against social imperfections. And we favor social reform rather than social revolution because violence is wasteful and ends up by eroding everyone's morals. (After all, wars, whether civil or international, hot or cold, are fought to be won, and when the supreme goal is victory all means are deemed to be acceptable. Violence breeds immorality, which in turn begets violence: this is the worst vicious circle imaginable.) Conformism and revolution are only acceptable if the issue is one of life and death: the former when no revolt is possible, the latter when only revolt may make survival possible.

All social reform brings about some moral reform. The converse is false: social change cannot be brought about through moral regeneration. True, every one of us can and should improve his morals, hence his social behavior, but there is only so much that the isolated individual can do to change society. When a moral problem is too big for the individual, he will tend not to think about it for feeling impotent to tackle it. Social issues are so big that they can only be tackled by group action. Think of mass murder, poverty, and environmental degradation. Every one of us ought to refrain from killing, exploiting and littering, but our individual contributions will not solve or even palliate any social problems.

It is not that we must wait for the New Jerusalem to be born until a new morality can be adopted, and in the meantime we may go on behaving badly. On the contrary, those who can glimpse a better morality than the one practiced by the wolves or the sheep around us, ought to start practicing it and ought to band together to work for a better social and moral order. In the process we shall become better individuals and shall set an example to others. Moralizing without doing is ineffective and hypocritical. But doing without regard to moral values can be harmful. If we are in a hurry to do good we may overlook the right. Therefore we should think twice before acting.

Our preference for social reform over social revolution is based on the right to life and on a prudential consideration, namely that, because a social revolution mobilizes huge masses of people, it can hardly be controlled and therefore it is bound to take unpredictable turns and to have undesired consequences. For these moral and practical reasons political violence should only be resorted to when directed against a merciless power intent on slaughtering or starving an entire social group.

However, our preference for social reform over social revolution does not amount to endorsing the policy of piecemeal social engineering recommended by liberals from Bentham and Mill to Popper and Rawls. Since every human society is composed by persons with biological needs, who live in an environment and are organized into economic, political and cultural systems, any realistic social reform is to be global or systemic, not one-sided or partial. Moreover reforming a single aspect of social life may do more harm than good. For instance, introducing sanitation in a community without at the same time improving the economy may lead to starvation by way of overpopulation; and introducing high yield grains without at the same time reforming the system of land tenure will ruin the poorest peasants. Social reforms must be global and participatory as well as gradual and designed with the help of the best available knowledge if they are to be efficient and lasting.

To adopt a global or systemic view of social reform entails rejecting all of the policies of partial social development advocated nowadays. We reject pure environmentalism because the economy must be kept going lest we all starve; pure biologism because we cannot keep in good health without a clean environment and an adequate income; pure economicism because economic prosperity is worthless unless we enjoy good health and can make use of our income; pure politicism because freedom and participation are pointless if we are sick, destitute, or ignorant; and pure culturalism because the production and consumption of cultural goods take health, economic means, and a modicum of freedom and leisure.

To put it in positive terms: Progressive social reform, that is, social development, is at the same time environmental, biological, economic, political, and cultural: See Figure 11.1. In keeping with our pentagonal view of social development, we shall expound our social philosophy in five parts.

1. ENVIRONMENTAL

Ever since the Industrial Revolution started two centuries ago, and particularly since the beginning of WWII, industry has been plundering

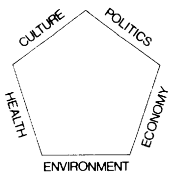


Fig. 11.1 The five aspects of social development: environmental, biological, economic, political, and cultural.

and contaminating our habitat. If the exploitation of the non-renewable resources were to continue at the present rate, very soon there would be no oil left, and many metals indispensable to industry would become extremely scarce (Brown *et al.* 1987). If toxic wastes, such as trace metals, continue to be dumped, our grandchildren will be malformed, severely sick, or short-lived (Nriagu & Pacyna 1988). Unless the deforestation and desertification processes are halted, half of the extant biospecies will become extinct before the 21st century is over (May 1988). This is where the *laissez-faire* — or rather *laissez-défaire* — environmental policy is leading: to the quick transformation of our planet into a place uninhabitable by human beings, or at least by civilized and healthy ones.

What is to be done? There are two possible solutions: Romantic conservationism or scientific environmental protection. The *romantic conservationists*, or radical "greens", believe that industry disturbs the ecological balance, which they regard as static in nature. Hence they preach the replacement of modern industry with traditional craftmanship, and consequently the return to primitive lifestyles. But their assumption that natural ecosystems stay in equilibrium is not true: ecosystems evolve, so there is no point in attempting to conserve them. (See Küpper 1982.) And deindustrialization would cause more than the destruction of civilization: it would bring about a hecatomb. (Think how long you could survive without any industrial goods or services.) Their romantic project is unfeasible for it would never win public support, particularly in the underdeveloped countries, where the problem is not overindustrialization but its opposite. Nor is it morally justifiable, because it demands sacrifice without compensation — except for the bald eagle and the piranha. Luckily there is a rational alternative, namely scientific environmental protection.

The *scientific environmentalist's* diagnosis of the environmental crisis is roughly the same as that of his romantic fellow environmentalist. We are causing the degradation of the biosphere, wiping out numerous biospecies, and soiling our own nest to the point of having turned our species into a highly endangered one. Some of us are consuming far beyond our means while others are reproducing beyond our planet's carrying capacity. However, the scientific environmentalist's prescription is not to reject technology and turn the social clock back, but to adapt technology to our real needs and plan social development on a world scale so as to protect our habitat. Let us use our gray matter to protect our green world.

The idea of the scientific environmentalist is to combine the scientific and global management of the biosphere with the social reforms necessary to ensure the success of environmental management. The reasons for favoring such a combination are that (*a*) purely environmental protection measures might only worsen the poverty that afflicts at least four-fifths of the world population, and (*b*) social reforms unaccompanied by an energetic world-wide environmental protection action would end up in a good society of the dead.

In other words, the goal of the scientific or green-gray environmentalist is to eliminate the conflict between development and environmental protection which has characterized the socialist as well as the capitalist societies heretofore. Shorter: the proposal is to adopt what may be called *ecosociodevelopment*, i.e. integral (not just economic, cultural or political) development together with environmental management. (See e.g. Thibodeau & Field Eds. 1984, WCED 1987, Bunge 1989.) This policy opposes both of the rival slogans, *laissez faire la nature* and *laissez faire l'industrie.* It involves transforming (not conserving) nature so that it may coexist with civilization, and transforming (not destroying) the latter so that it may last. See Figure 11.2.

Designing and implementing ecosociodevelopment involves much more than such purely environmental protection measures as rational waste management, reforestation, and desert reclamation. To be effective and lasting, the reform must cover nearly every aspect of social life everywhere and it must win the support of all but those who are sick with economic or political greed. The reform must include

(1) a gradual but eventually total nuclear, chemical and biological disarmament, because the arms race is depleting the non-renewable resources, it involves highly polluting industries, and it risks culminating in a nuclear war followed by the nuclear winter, which could wipe out all life on earth;

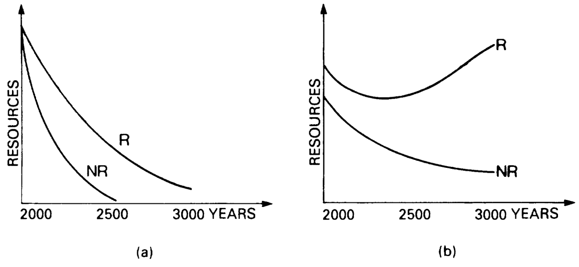


Fig. 11.2. (*a*) Depletion of renewable (*R*) and non-renewable (*NR*) resources if they continue to be exploited at the present rate. (*b*) Rational management of world resources: increase of renewable (*R*) resources up to the necessary level, and decrease of the rate of consumption of non-renewable resources (*NR*) as an effect of restrictions on exploitation, substitution of fission energy for fossil fuels, wood and plastics for metals, recycling, etc. Arbitrary scales. (Redrawn from Bunge 1989.)

(2) the replacement of international confrontation with negotiation and cooperation regulated by the United Nations and, eventually, by a world government, to make war not just unlawful but also unprofitable;

(3) the establishment of an international environmental management agency with authority to prevent and remedy environmental catastrophes the world over;

(4) the gradual phasing out of all the nuclear energy plants, until we discover how to dispose of nuclear wastes and how to prevent the plutonium generated in some of them from being employed to manufacture nuclear bombs;

(5) massive investment in research, by international teams, into alternative energy sources such as fusion;

(6) the redesign of all machines and industrial processes with the aims of (*a*) increasing their efficiency, (*b*) reducing the production of pollutants, and (*c*) treating (recycling or breaking down) the waste, wherever possible in place;

(7) upgrading phytotechnology and agricultural engineering to improve agricultural yields without eroding or otherwise impoverishing the soil — or the farmers;

(8) a strict control of the use of artificial fertilizers, insecticides, herbicides and defoliants;

(9) phasing out the use of oil as a fuel, allowing it to be used exclusively by the petrochemical industry for the manufacture of plastics, synthetic fibres, and the like;

(10) internationalizing the non-renewable resources and placing them under an international agency empowered to lease them to private firms or cooperatives bound to observe the exploitation quotas and the efficiency norms set up by the agency;

(11) the extension and improvement of the mass transportation systems, with a view to reducing drastically the number of cars — some of the worst environmental offenders;

(12) adapting technology and industry to meeting only the basic needs and legitimate aspirations of people;

(13) the banning of all publicity aimed solely at increasing unnecessary consumption;

(14) upgrading the educational system on all levels and, in particular, intensifying ecological education and the education of consumer habits;

(15) popularizing family planning in order to decrease the rate of exploitation of natural resources — and of course in order to improve the general quality of life.

A peaceful world-wide social reform so radical and many-sided as the one sketched above cannot be achieved overnight and without a concomitant change in value systems and morals. It will have to be accompanied by the adoption of a value system and a morality in harmony with ecosociodevelopment. We submit that the views on values and morals proposed in this book are the suitable ones, because they are based on the recognition of the basic needs and legitimate aspirations of all human beings. Indeed, if life is the supreme value, enjoying it the supreme right, and helping others live the supreme duty, it stands to reason that we must start by securing a suitable place to live in. And if this place is to be kept inhabitable, we must redesign social behavior accordingly, for managing anything, in particular environmental items, boils down to managing people. (Recall Ch. 10, Sect. 2.5.)

The very idea of ecosociodevelopment will be found revolting by all the moral philosophers who preach the sanctity of self-interest: moral nihilists, rational egoists, libertarians, contractualists, and negative utilitarians (Ch. 7, Sect. 2). It will also be rejected by the political scientists and philosophers who believe that liberty — in particular the liberty of doing as we please with nature — comes before everything else, even before welfare. Likewise it would be rejected by the economists who equate development with economic growth, wax enthusiastic over the power of technology-based industry to subdue nature, and oppose any attempts to limit industrial growth and adjust technology and industry to the basic human needs and legitimate aspirations. After all, we have come this far in the process of destroying our habitat by taking advantage of the alleged boundless freedom to exploit the natural resources. It only takes the *modus tollens* to realize that, given the environmental mess we are in, we must reject the premise that we are free to go on plundering and soiling our habitat. Since such freedom has got to be curtailed, it had better be done rationally, i.e. as part of a comprehensive plan to save the Earth from and for ourselves and our offspring.

Some of those who believe that everyone has the right to pursue what he takes to be the good life may suspect that the ecosociodevelopment we favor involves adopting an ascetic lifestyle. Not so. Rather, unless we replace the current prodigality of the few with the frugality of all we shall condemn our posterity to monasticism and eventual extinction. Far from preaching the joyless life, we repeat our slogan *Enjoy life and help live*, and add the following unavoidable platitudes: (*a*) at present most people do not have the means to enjoy life, and many of those who do have them mistake the good life for the ability to buy whatever they fancy; (*b*) unless we alter some of our values and learn to administer wisely our resources, we shall rob our offspring of their inheritance.

Even some of those who agree that something ought to be done to save our planet may believe that nothing can be done: that "people" will continue to squander our common heritage until it will be too late to save what may be left of it. However, the recent agreements on nuclear arms limitation, the ban on the production of sprays and plastics that are destroying the ozone layer, the successful vigilance of pollution levels in the Mediterranean and in the Rhine, and the internationalization of Antarctica, show that it is possible to reach consensus on certain environmental measures. Every large scale man-made environmental catastrophe heightens the public awareness and puts pressure on governments and industries: though admittedly slowly, we are making progress.

Still, it may be objected that ecosociodevelopment on a planetary scale is unfeasible as long as the world remains divided into antagonistic blocs, and every society is split into mutually opposed classes. But this is precisely the point: that in the long run the environmental crisis is far more serious than any international or social conflict for, without breathable air, drinkable water, and uncontaminated food in sufficient quantities, we will not survive and consequently there will not even be political or economic issues to fight over. Besides, pollution and the depletion of non-renewable resources are equally serious in the socialist and capitalist blocs. And sometimes rival social groups, such as the workers and managers of polluting industrial plants, or the slash-and-burn farmers and ranchers in the Amazon, disagree on everything except the disregard for the environment and consequently for the future. To sum up, the environmental issue is prior to every other contemporary issue except for its twin, the nuclear arms race. However, neither of these two monsters can be tamed without radical world-wide social reforms, for the management of any resource boils down to the management of people.

To conclude, we must protect our planet, but not at the cost of social development. The alternative is neither environmental protection nor social development, but either a continuation of the present course towards ultimate environmental catastrophe, or the improvement in the quality of life for everyone in and through ecosociodevelopment. The order is tall but the stakes are high. (See the Declaration of The Hague, of March 11, 1989.)

2. BIOLOGICAL

Most moral philosophers interested in bioethics have focused their attention on such emotionally and ideologically charged issues as abortion, euthanasia, and suicide. Though genuine, these problems affect only a small fraction of humankind. The most serious bioethical problems are those of starvation, malnutrition, sickness, and child neglect, for they affect billions of people and are likely to grow before they are solved. We submit that the root causes of these problems are unbridled population growth, environmental degradation, economic inequality, political power exerted to maintain inequality, and military expenditures. Wherever the population is commensurate with the resources, wealth is distributed equitably, and the government protects the weak and the environment, those problems hardly arise. Let us start with overpopulation.

The world population has increased to about ten times its value at the beginning of modern times. If the present growth rate of 1.7 per cent per year were to remain constant, the world population could double in less than thirty years — unless it were to be reduced by famines, plagues, and wars. Extrapolating present trends, demographers project that around the year 2010 India's population will be stabilized at 1,700 million, Nigeria at 530, Indonesia at 370, and Pakistan, Bangladesh and Brazil at 300 million each. (See e.g. Brown *et al.* 1987.) Reproductive success breeds population excess, which in turn begets misery and worse.

The effects of overpopulation, particularly when accompanied by economic inequality, are poverty, unemployment, subemployment, rapid environmental degradation, intensification of conflicts over the control of natural resources, and the growth of governments, which are increasingly unable or unwilling to provide social services, in particular health care and education. In the so-called developing countries — which are actually declining — roughly half of the population is under 15 years of age, most children do not get appropriate nourishment and health care, they cannot complete their elementary education, are often unwanted — hence neglected or even homeless — and when they reach maturity they seldom find jobs. In short, as the world population continues to grow, the quality of life of most people goes down, and the prospects of our posterity grow bleaker.

Thousands of years ago, when there were few people and their life expectancy was short, promoting human fertility was the prudent and moral thing to do. In the present world, given that overpopulation is a root of so much evil, it is irrational and perverse to preach reproductive freedom. It is irrational because our ecological niche is nearly filled to capacity, and perverse because it condemns millions of children to neglect, hunger, early death, and ignorance, and hundreds of millions of adults to squalor and despair. Reproduction is no longer a human right: human fertility must be controlled in the interest of the survival of humankind. (Note the paradox: To remain fit we must check our Darwinian fitness.)

There are three ways of checking population growth: letting pathogens take their toll, waging genocidal wars, and birth control. The Black Death of the 14th century exemplifies the first way. It killed an estimated 25 million people — one-fourth of the European population — and in some countries it mowed down up to three-quarters of the population. But do not count on epidemics to solve any problems. The influenza epidemics of 1918 to 1919, which made almost as many victims as the plague in the 14th century, only added to the misery caused by the world war that preceded it. As for genocide, it was more effective in ancient times, when most ethnic groups were comparatively small.

The only rational way to control population growth is birth control. This may be done in three ways: (*a*) by infanticide and abortion, (*b*) by raising the standard of living and the level of education, and (*c*) by adopting and enforcing strict planned parenthood legislation. Infanticide is of course immoral, and abortion is traumatic on top of being unnecessary given the availability of contraceptive methods. The second method is indirect and it has worked well in most industrialized countries, which have succeeded in reducing their birth rates; but of course it is inapplicable in poor societies, which are the most numerous. These should resort to the third or direct method, namely compulsory family planning. The only objection to this method is theological and even so not cogent because, unlike the case of abortion, no embryonic human life is at stake when fertilization is prevented. Preaching fertility to people who do not earn enough to feed themselves is downright immoral.

Let us wrestle briefly with a handful of other bioethical issues which, though serious, pale by comparison with that of overpopulation. Our first question is whether everyone is entitled to state-supported health care. Our answer is affirmative because (*a*) everyone has the right to live, (*b*) only the state can embark on massive health care programs such as the construction of sewers and compulsory inoculation, and (*c*) wherever there are disadvantaged people, only a national health program can take care of their health. Right-wingers oppose such programs in the name of liberty. (Recall Ch. 7, Sect. 2.3.) But minimally intelligent people support them because they understand that their own health is at stake: it is impossible to enjoy good health when surrounded by sick people. In short, prudence and morals speak in favor of state-supported health care.

Our second question is whether the state has the right to protect its subjects from themselves. Mill and other libertarians have denied that it has this right, and have defended everyone's right to drug, drink, or smoke himself to death, or even to mutilate himself. Against this claim it may be objected that (*a*) if we wish virtue to prevail we must not set the example of vice; (*b*) certain vices, particularly the consumption of alcohol and hard drugs, induce antisocial, often criminal, behavior; (*c*) the treatment of the medical and social problems caused by indulgence in such vices is usually entrusted to public institutions, particularly hospitals and jails, hence at the public's expense; and (*d*) the production of the means required to satisfy these vices places an unnecessary burden on the soil. The only problems, then, are of a practical nature, namely how best to prevent and cure drug addiction. The informed opinion is that education is the best prevention, and that addiction can be treated successfully by a combination of medication with behavior therapy. Cut down the demand, and the offer will decline automatically together with the price. If on the other hand all the effort is directed against drug production and trade, the net effect is that the drug business is controlled by organized crime. In conclusion, the state has the right and the duty to protect the individual from himself, for by so doing it protects relatives, neighbors, and the population at large.

Our third question concerns genetic engineering. This technology raises at least two bioethical problems: the production of new biospecies that might endanger human health or the environment, and the production to specification of new varieties of human being. The former can be avoided by adopting and enforcing restrictions similar to those that govern the manufacture of new pharmaceutical products. The opposition to the production of new biospecies or varieties, such as frost-resistant or drought-resistant crops, is just a variety of fanaticism. If such fanatics had had their say fifteen millennia ago, when artificial selection was first practiced, we would still be food gatherers rather than producers.

As for human genetic engineering, so far its aim has been to prevent or cure hereditary diseases, and this is all to the good. But there is also the possibility that human beings may be manufactured almost to order, either to increase uniformity or to produce special varieties, e.g. of obedient and dumb manual workers or soldiers, as envisaged by Aldous Huxley in *Brave New World.* Such technology would be morally and practicable undesirable. Morally because it would enhance social inequalities to benefit the privileged. And it would be practically undesirable because we ought to let chance play its role in shaping populations where every desirable trait be distributed normally, i.e. on a bell-shaped cure. Human genetic engineering is desirable only to reduce the number of severely handicapped individuals. Our present degree of social and moral development does not warrant looking indifferently at the prospect of using genetics to skew the normal distribution of, say, physical strength or intelligence.

Our last question concerns animal rights: Should we treat other animals on a par with humans, particularly with regard to inflicting pain on them? There are moral and practical reasons for sharing some, though not all, of the tenets of the animal rights movement (for which see Singer 1975). The moral reasons are that (*a*) most of the animals we have commerce with are capable of suffering and (*b*) cruelty to animals makes us more inclined to be cruel to fellow human beings: being kind to animals is part of our moral education — as long as it is not accompanied by indifference to human suffering.

However, the practical reasons for refraining from harming, and particularly killing, other animals are perhaps even more compelling than the moral reason. One is that, as all ecologists know, human survival depends on that of millions of other biospecies, every one of which participates in some food-web or other. Another is that raising cattle and fowl for meat is economic lunacy. Remember that the production of 1 kg of beef protein requires the expenditure of about 10 kg of protein contained in hay, grain, soya beans, fishmeal, etc. We need meat substitutes, and food technology will invent them if demand for them increases. As for milk and eggs, genetic engineering can prevent the birth of more males of the domestic species than are necessary to perpetuate these species.

As for animal experimentation, we cannot endorse the view of the so-called animal liberation movement. To raid biological laboratories and set the experimental animals "free", i.e. at the mercy of an environment alien to them, is insane. Biological experimentation is indispensable not only to increase scientific knowledge but also to improve human health and to improve the welfare of domestic animals. (The latter is, precisely, a goal of applied experimental ethology.) What we must do is to refrain from causing unnecessary pain and from performing mindless, hence useless, experiments just "to see what happens" when harmful stimuli are applied to an animal. Such precautions should be taken particularly when using human subjects.

Finally, the norm that we should be kind to animals does not entail that we should not interfere with wildlife. On the contrary, we must regard wildlife management as part of environmental management, because some wild biopopulations are about to become extinct, while others are threatened by overpopulation. The first problem has been solved successfully in the case of a few species that have captured the imagination of conservationists. The overpopulation problem has traditionally been solved by trappers and hunters, but this solution is neither rational (because it may lead to extinction) nor humane. A more humane method is contraception, which has been employed to check the pigeon population in cities. (More on bioethics in Ch. 8, Sect. 2.2.)

3. ECONOMIC

Our economic philosophy rests on the ethical principle that everyone has the right to meet his basic needs and fulfil his legitimate aspirations, as well as the duty to help others exercise the same right (Ch. 4, Sect. 1). Anyone who admits this principle is committed to adopting

NORM 11.1 The one and only morally legitimate function of the economy is to help people meet their basic needs and fulfil their legitimate aspirations.

This rule is not as simple as it may look. To begin with, it entails a condemnation of human exploitation and of the exclusive pursuit of profit. Indeed, it suggests advocating distributive or social justice of some kind. Moreover the norm leads to condemning all wasteful economic activities, as well as those which involve the production of purely destructive or even merely superfluous commodities. Finally, the norm renders it nonsensical to claim that the economy is in good shape when there are destitute people, or that it is in bad shape when there are none.

Let us glance at some of the problems posed by the adoption of Norm 11.1. A first point to be noted is that mainstream economic theory will not help implement the norm for the following reasons. Firstly, that theory is immoral for postulating that economic behavior is, or ought to be, "rational" in the sense of maximizing the expected utility of the individual agent regardless of the well-being of others. (For an indictment of egoistic views see Ch. 7, Sect. 2.) Besides being immoral the "rationality" (selfishness) postulate is imprudent, because a society of egoists would soon disintegrate unless held together by violence or deceit, neither of which is moral. Secondly, mainstream economics takes for granted that individuals have the right to own any natural resources. This is immoral because there is only so much of any natural resource, and not enough for everyone to become a land-owner, mine-owner, or forest-owner. Thirdly, the theory presupposes that whoever has the wherewithal is entitled to exploit labor. For these reasons we cannot count on mainstream economic theory to design a just economic theory. (Nor does the theory describe adequately any of the existing social orders: Recall Vol. 7, Ch. 4, Sect. 5.2.)

A second point concerns the myth, shared by orthodox Marxists as well as by apologists for capitalism, of indefinite economic growth. This assumes either that the world's resources are inexhaustible, which is false, or that it does not matter if we disinherit our posterity through the overexploitation of the environment, which is immoral. Economic growth must be limited if we are to bequeath a livable planet to our descendents: Recall Sect. 1. But it must not be limited uniformly throughout the world, for at present most people in the world are poor. It would be immoral to ask the poor to refrain from aiming at economic development. Complying with Norm 11.1 requires aiming at a worldwide economic level situated between subsistence and abundance.

Now, classical capitalism is inherently expansive: it conquers new internal and foreign markets supplying new consumers and offering new investment opportunities. The reason for this expansiveness is that, through the use of continually renewing technologies and the abuse of the environment, a capitalist economy increases its productivity and accumulates capital. Unless the excess commodities find consumers, and the excess capital can be invested, the rate of profit is bound to decrease, and unemployment to increase. In turn, the decline in the rate of profit discourages investment in industry, which consequently becomes technologically obsolete and thus less productive; and unemployment shrinks the internal market and forces the government to raise taxes in order to maintain social services. The overall outcome is at first stagnation, later on decline. Such decay is unavoidable if half of the engineers, and half of the richest industrial companies in a country, are devoted to the development and production of weapons: See Bellon & Niosi (1988).

Since economic expansion is not intrinsically valuable, but should only be a means to meet public demands, and it should be controlled to prevent the degradation of the environment, classical capitalism is not the ticket. Nor is it classical or statist socialism, because it is inefficient and authoritarian. It is inefficient because it suffocates individual initiative, dilutes personal responsibility, and underrates incentive; and it is authoritarian, hence alienating, for involving the subordination of the economy to the party. The consequences are well known: shortage, waste, low quality, worker dissatisfaction, and parasitism.

If we wish to build a just and viable economy we must move beyond both classical capitalism and classical socialism, to a sort of synthesis of the two without their disadvantages. Actually the convergence of capitalism and socialism is already under way: witness welfare capitalism and the "new capitalism" on the one hand, and cooperativism and the *perestroika* (restructuring) of statist socialism on the other. (See e.g. Nienhaus 1984, Halal 1986, Stojanović 1988.) In fact, welfare capitalism redistributes part of the national income, and cooperativism eliminates exploitation altogether; the "new capitalism" gives more of a say to workers and consumers, and *perestroika* is an essay in market socialism.

To be sure, we are only at the beginning of the process, and there are many hurdles to be overcome. The welfare (actually relief) state is in trouble in some countries, not only because of the hostility of short-sighted pro-capitalists but also because of the military build-up, high unemployment rates and the heavy overhead costs of relief administration, not to mention corruption. As for cooperatives, so far they have operated either in a hostile environment, or under the thumb of politicians and with little technological (particularly managerial) inputs. The "new capitalism" has yet to implement a more intensive participation of its work force and consumers, and market socialism calls for the dismantling of unprofitable enterprises and a vigorous shake-up of personnel. And all, whether reformed capitalists or reformed socialists, have got to clean up the fouled environment and put an end to international confrontation.

The solution to these difficulties lies in furthering the convergence process, building economies which operate on the principle of competitive cooperation and subject to Norm 11.1. We submit that, from a practical as well as from a moral point of view, the way to achieve this goal is to combine cooperative ownership with technodemocratic management and competition, the whole regulated (but not ruled) by the state to safeguard the public interest. Let us see why.

Cooperative ownership has, at least in theory, the following advantages over alternative types of ownership: equitable distribution of wealth (social justice), personal concern for the success of the enterprise (hence efficiency), encouragement of individual initiative (hence openness to innovation), democratic control of management (hence work satisfaction), decrease in conflict intensity (hence decreased disruption), and personal satisfaction derived from working in a friendly environment. (See e.g. Dahl 1985.)

The advantages of competition — challenge, productivity, concern for quality, openness to innovation, and minimization of parasitism — are secured by allowing for the multiplication of cooperatives in nearly every sector of the economy. However, since unbridled competition can be destructive of both firms and the environment, it should be controlled by a neutral party, the technodemocratic state. (It goes without saying that the state is also necessary to provide essential social services that no cooperative and no corporation can provide to every person that needs them.) Without such neutral control serious inequalities may develop — e.g. between the returns of a bakery and those of an academic publishing firm. Such inequalities in profits — though not those occurring within a given sector — may be diminished by taxation and redistribution of the tax revenues.

The state would intervene in the economy only to have Norm 11.1 observed. In particular it would intervene to (*a*) support the production and commercialization of low return but socially useful goods and services, (*b*) prevent the formation of monopolies and monopsonies, except in the military sector; (*c*) prevent the production of dangerous things, such as hard drugs and handguns, and the offer of dangerous services, such as medical quackery and mercenary soldiering; (*d*) enforce tight controls on the exploitation of natural resources; and (*e*) monopolize the manufacture and use of weapons. The market resulting from such state intervention in the operation of cooperative enterprises would combine the desirable traits of capitalism and socialism. In fact it would meet the basic needs and legitimate aspirations of everyone, it would ensure efficiency in the allocation of resources (e.g. by leasing land tracts to the most productive cooperatives), encourage the production of a variety of high quality commodities, and make ample room for initiative and innovation.

Table 11.1 exhibits some of the advantages and disadvantages of the various possible types of industrial economy. The most obvious virtue of competitive cooperativism *cum* flexible overall planning is that it combines social justice with democracy. Aristotle, who wrote that "poverty is the parent of revolution and crime" (*Politics* Bk. II, Ch. 6, 1265), might have approved of it. Even Smith, who deplored that "Wherever there is great poverty, there is great inequality" (1776 Bk. V, Ch. I, Pt. II), might have preferred it to corporate capitalism, with its tendency to monopoly, which he condemned for killing free enterprise. And it would certainly have met with the approval of Mill (1852 Bk. II, Ch. I, Sect. 3, and Bk. IV, Ch. VII, Sect. 6).

Competitive cooperativism is part of integral or social democracy but it does not involve radical egalitarianism. On the contrary, it

TABLE 11.1. Economic orders.

Type of economy	Efficiency	Fairness to worker	Fairness to consumer	Worker incentive	Management incentive	Worker satisfaction	Management satisfaction
Classical capitalism	½	0	½	½	1	0	1
Welfare capitalism	½	½	½	½	1	½	½
Welfare capitalism *cum* coordination	1	½	½	½	½	½	1
Classical (statist) socialism	0	½	0	½	½	½	½
Competitive cooperativism	½	1	½	1	1	1	1
Competitive cooperativism *cum* coordination	1	1	1	1	1	1	1

involves income differentials adjusted to needs and deserts: i.e., it goes hand in hand with qualified egalitarianism (Ch. 6, Sec. 2.2), according to which persons with similar merits deserve to occupy the same rank. This form of egalitarianism involves (*a*) the recognition of merit, as well as of the special responsibilities dependent on special abilities, and (*b*) the principle that the public is best served by competent and satisfied workers, managers, and public servants. In short, competitive cooperativism is meritocratic: it admits all and only the inequalities likely to favor the majority. By the same token it admits neither privileges, i.e. rights not earned ("entitlements"), nor underprivileges, i.e. burdens not deserved. Which takes us to the heart of the matter, namely social justice.

The scholarly literature on social or distributive justice deals almost exclusively with the distribution of benefits, in particular goods such as money and services such as health care. But in real life we judge justice and injustice not only in terms of benefits received but also in terms of services rendered. Shorter: Social justice is a matter of sharing benefits and burdens. In a just society benefits and burdens are distributed equitably, i.e. everyone's burden is balanced by his benefit. More precisely, in line with Definition 6.5 we propose

DEFINITION 11.1 A society is *internally just* $=_{df}$ every member of the society

(i) receives what she requires to meet her basic needs;

(ii) can earn, by doing socially useful work, what she requires to satisfy her legitimate aspirations;

(iii) fulfils the obligations assigned her by her family, workplace, and social circle(s) — assignments which, in the case of adults, are made by mutual agreement;

(iv) is free to satisfy her legitimate aspirations and to pursue those inclinations that are not antisocial;

(v) is free to work for points (i) to (iv) above, by herself or in association with others.

Note the following points. Firstly, social justice is a systemic property and, more particularly, a social one. But of course it emerges from the interplay of individuals. Prosocial behavior promotes social justice while antisocial behavior thwarts it. Secondly, a society with a subsistence economy, or one with an industrial economy that tolerates unemployment, is not socially just even if it provides a subsistence income, because such a society fails to satisfy conditions (ii) and (iv). Thirdly, according to clause (iii) social justice involves democracy in the workplace, i.e. self-management. Fourthly, clause (iv) protects the right to pursue hobbies and to engage in idiosyncratic behavior as long as neither harms others. Fifthly, clause (v) makes it plain that social justice cannot be maintained without a modicum of freedom, in particular freedom of speech and association.

The above definition is ineffective unless supplemented by a precise formula for the distribution of income. We propose the one contained in

NORM 11.2 The just income of an individual equals the cost N of meeting her basic needs plus the market value W of the socially useful work that she performs: $I = N + W$.

In other words, in a just society you receive what you need to survive plus what you earn by honest means. The formula encourages socially useful behavior and discourages antisocial behavior. It combines welfarism with meritocracy. Unlike the fixed wages option, it rewards skill and effort. And, unlike the "what you can earn" option, it makes provision for the basic needs of the people who cannot work. The formula quantitates the principle: *To each according to her needs and deeds, from each according to her abilities and responsibilities.*

However, our formula has a catch: Who determines what "socially useful work" is? Are the activities of punk noisicians, athletes and vulgar TV comedians socially useful just because there is a demand for them? We suggest that this problem will be solved in practice, though not government decree, as culture evolves. Vulgar entertainment is part of a commercialized culture and it should wane along with the rise in cultural level made possible by social justice. More in Sect. 5.

Definition 11.1 concerns the internal social order not the international one. Now, certain nations, notably Great Britain, France, and the U.S.A., have attained a certain level of social justice partly at the expense of the Third World, the natural and human resources of which they have exploited exclusively in their own benefit. Therefore we also need

DEFINITION 11.2 (i) A society is *externally just* if and only if it does not hinder the development of other societies. (ii) A society is *just* if and only if it is both internally and externally just.

Internal (or social) and external (or international) justice are desirable not only for moral reasons but also prudential ones, namely because they are the only rational and lasting means of preventing civil strife and international conflict.

Finally, what about property rights in a just society? This is a legal problem not a moral one because, in order to satisfy our basic needs and legitimate aspirations, we only have to have *access* to certain things without necessarily owning them. We do not need to own a person in order to marry her, a mountain in order to enjoy its view, a school in order to learn, or a bakery in order to eat bread. Property implies possibility of use, but the converse is false. Moreover there are well know objections to the private ownership of factors of production: (*a*) it causes the division of society into the propertied and the rest; (*b*) it jeopardizes the right to life of the propertyless; and (*c*) it is the major cause of uncounted conflicts and crimes, from litigation to war. For these reasons we propose a principle of restricted property rights:

NORM 11.3 In a just society everyone may own whatever she requires to meet her basic needs and legitimate aspirations, as well of those of her dependents, and nobody may own more than this.

4. POLITICAL

In any society politics is action aimed at influencing or controlling the specific functions of the state — municipal, national, or other. Remember that a specific function of a thing of a given kind is a process that only things of that kind can go through. (See Vol. 4, Definition 3.8.) For instance justice, public security, defense, and the redistribution of national income are specific functions of the state, because no other organization can perform them. On the other hand it is debatable whether the execution of public works and the management of public transportation, or even public education and the mail service, are specific functions of the state. Indeed, since they can also be performed by business corporations or by cooperatives, it is up to the citizenry to favor one or the other form of management of such public services. Many debates in contemporary political and legal philosophy are precisely about this point.

Political and legal philosophy are centrally concerned with the *legitimate* functions of the state, whether or not they are specific, as well as with the rights and duties, political or otherwise, of the individual, in particular the citizen. Unlike political science and jurisprudence, which are mainly descriptive and explanatory, political philosophy is mainly evaluative — in particular critical — and normative. It is concerned with what is good or bad, right or wrong in the polity. For example, whereas a political scientist's major interest may be to ascertain the efficiency and stability of a political regime, a political philosopher will be interested mainly in ascertaining whether that regime facilitates the pursuit of well-being, or whether certain reforms would improve welfare.

We can only touch on a handful of topics in political philosophy, starting with the legitimate functions of government. A good government ought to make it possible for everyone to practice our supreme moral norm, i.e. to enjoy life and help live. (Note the cautious 'possible'. Well-being must be felt, and solidarity practiced: neither can be dispensed.) Needless to say, none of the existing governments complies fully with this condition. Worse, most of them oppress and milk their subjects, legalize exploitation (in some cases on an international scale), take part in the arms race, and do nothing to protect the environment. Even the relief state, though by far preferable to its predecessor, is imperfect for it only compensates to some extent for social injustice. Besides, by interposing a bureaucrat between any two individuals, it shrinks personal responsibility and undermines mutual help, which, jointly with participation, constitutes the mortar that holds social groups together. Still, the welfare (or rather relief) state does go some way in the right direction. Hence any attempt to dismantle it without solving the social problems that prompted its creation is morally objectionable.

We agree with Aristotle's view that "a state exists for the sake of a good life" (*Politics* Bk. III, Ch. 9, 1279). Hence with Jefferson (1809) when he wrote that "The care of human life and happiness, and not their destruction, is the first and only legitimate object of good government". More precisely, we posit

NORM 11.4 The legitimate functions of the state are (i) to protect the basic human rights and enforce the corresponding duties (Ch. 4, Sect. 1.2); (ii) to facilitate the fulfilment of everyone's legitimate aspirations (Ch. 1, Sect. 3.1); (iii) to facilitate and coordinate the activities of the various non-governmental economic and cultural organizations that operate in the public interest; (iv) to manage the public goods; and (v) to deal and cooperate with other governments.

Probably the most problematic concept in this norm is that of a public good. Let us define it before discussing it:

DEFINITION 11.3 A *public good* is a thing or service that

(i) is necessary for the welfare of the members of a social group, possibly for all the members of society;

(ii) is accessible to all the members of society — though not necessarily for free; and

(iii) cannot be lawfully appropriated by any individual or group other than the organization that provides it.

Note that this definition makes no reference to ownership. In particular it does not contain the liberal view that all public goods should be owned by the state: this is a thesis, hence it ought not to be part of a definition. This thesis has been rejected by the anarchists (e.g. Kropotkin 1902, Taylor 1976), who claim that voluntary cooperation can provide the same public goods without any need for government — which is probably true for a Neolithic village, but certainly false for any modern society. At the other extreme, contemporary conservatives claim that all public goods, with the exception of the courts of law, the internal security forces and defense, can be provided by business corporations to the tax-payer's advantage — provided of course he can afford to pay for them. But any businessman and any social scientist knows that the market cannot supply such goods as environmental protection, city planning, traffic safety, consumer protection, public health, compulsory elementary education, and much of science, the humanities, or the arts. Only the state or state-supported cooperatives can provide such services because, though necessary for the welfare of all, they are not marketable: they happen to be *public* goods provided for public service not private profit. Consequently to preach the privatization of all or nearly all public goods is no less than preaching the dismantling of civilization.

The variety and quantity of public goods have increased enormously in all industrial societies in the course of this century. One result has been an unprecedented increase in welfare, another a frightening increase in government size. Granting that Big Government is wasteful, suffocates initiative, threatens freedom, leads to abdicating personal responsibility, and substitutes charity for solidarity, how can we shrink it without destroying civilization? The problem of government size may be analyzed into the subproblems of efficiency and morality. The efficiency problem is this. Who can manage most efficiently public services: the state, business corporations, or cooperatives? Experience shows that in some cases (e.g. European railway companies) the state does a superb job; in others (e.g. North American telephone companies) private business provides good service; and in still others (e.g. Argentinian microbus companies) cooperatives are remarkably efficient. Since the type of ownership seems to make no essential difference to efficiency, we are left with the moral problem.

The moral aspect of the question is this. Which form of public goods ownership harmonizes best with the ideal of social justice? Obviously private ownership does not, for a business corporation must refuse to serve anyone unable to pay for its services. The same holds for a cooperative, but not for the state. Were it not for the much-maligned welfare state, a large fraction of the population would have no access at all to health care, education, and other services. Moreover, given that some public services have got to be centralized, their privatization would condone business oligopolies or even monopolies, and therefore abusive price fixing. Consequently large sections of society would be deprived of public services — which, by definition, would cease to be public. Hence it is immoral to preach such privatization as a matter of principle. Does this commit us to favoring the state monopoly of all public services? Not at all.

Having submitted (Sect. 3) that cooperative ownership is the form of ownership best suited to a just society, we must favor *decentralizing the state and turning most of the public services over to cooperatives.* However, the state should retain the control of the quality of the services offered by the cooperatives, and it should use part of its revenues to correct pronounced inequalities between cooperatives as well as between regions. If society is to advance, the state cannot wither away: it must retain some traditional functions, such as justice; give up others, such as public transportation; and take on new ones, such as the management (though not the direct exploitation) of all the non-renewable resources, such as land and mineral deposits.

How about the poor, who might be unable to pay for the public services provided by cooperatives? This question presupposes that there will be poor people in a just society. But this is precisely the point: in a just society there *are* no poor people, except perhaps by choice. Consequently there is no need for state hand-outs except in cases of family tragedy or natural disaster.

In a just society there is no need for unemployment compensation because there is full employment. There is no need for state or private charity because everyone earns his living and that of his dependents.

The state bureaucracy is small because there is little need for public relief and most public services are provided by cooperatives. There is no standing army because international cooperation has been substituted for international conflict. Consequently the taxes are comparatively low and in any case they all go to pay for useful services. Finally in a just society there is no parasitism of the very poor or the very rich because there are neither and because, if an able-bodied adult were to refuse to work for a living, he would get barely enough to survive (Norm 11.2).

The political order accompanying a just economic order cannot be other than an integral technodemocracy (Ch. 10, Sect. 2.5). Democracy must be integral or social, not just political, economic or cultural, if everyone is to have an equal opportunity of getting access to all the resources of society. (Only the active partners in an organization participate in its management; the sleeping partners can be deceived or squeezed out. In a representative democracy the ordinary citizen is a sleeping partner.) And democracy must be technological, not amateur, if the management of the natural, economic and cultural resources is to be competent. (Note the difference between technocracy and technodemocracy: the former is the rule of experts, the latter that of the people in consultation with experts. Technocracy has never been tried, although bureaucrats are often called 'technocrats'. On the other hand technodemocracy has been tried on a small local scale and it has proved successful.)

Bureaucrats and technologists may dislike public participation in the management of the *res publica* because it ties their hands and wastes their time. But public participation is an effective deterrent of waste and corruption as well as a reminder to the public servant that he is supposed to be in the service of the people. Besides, the users and pecuniary supporters of a public service know from personal experience what its shortcomings are and they may come up with imaginative proposals undreamt of by indifferent key punchers. Finally, participation enhances responsibility, whereas marginality causes alienation — apathy, irresponsibility, despondency, or even rebellion.

What is the role of liberty in an integral democracy? It depends on the kind of liberty. Clearly, there must be freedom to work, but not to exploit; to pursue one's interests, but not at the expense of others; to indulge in harmless pleasure, but not in destructive or contagious vice; to compete, but not to destroy other people; to learn and teach, but not to force others to accept any beliefs; to run for elective office, but not to deceive one's constituency — and so on. The keys to integral democracy are equality, solidarity and participation. Freedom is necessary but insufficient, and one's liberties are limited by those of one's fellow human beings, near and far, alive or to be born.

How about legal paternalism? Does the state have the right and the duty to protect the individual against others and even against himself, e.g. by dissuading an individual from taking drugs? Legal paternalism, currently under attack on the part of libertarians and conservatives, is necessary to secure fairness and increase welfare (Garzón Valdés 1987). The only thing wrong with legal paternalism is its name, for advanced social legislation and social programs have often been the outcome of lengthy and bloody struggles. 'Legal fraternalism' is therefore a more appropiate name for it. What *is* morally and culturally objectionable is *moral* paternalism, i.e. the alleged power of the state to invade bedrooms, classrooms, libraries, laboratories and hospitals in an effort to enforce some obsolete moral code.

What about authority: is it legitimate for a democratic state to exert it? Authority is the power to force people to do what they do not want to. Hence it would seem that there is no room for authority in the democratic state. But of course no management is possible without some authority or other. Once a decision has been take democratically, those in charge of implementing it must have the requisite power, or the decision will remain dead letter. What we must avoid is not legitimate authority but illegitimate use of power, i.e. power abuse. And this is best avoided by decentralization, grass roots participation, freedom of speech, and freedom of association. Of course there must be some centralization in any state. But what characterizes the democratic state, as opposed to the authoritarian one, is central *coordination* rather than central *domination*: See Figure 11.3.

Ideally, in an integral democracy people would *want* to do their duty, for this would be in their own best interests. In such circumstances authority would be the power to delegate power and persuade people to do what is right. Any breach of authority would be regarded as either a mistake or an indication that the people in authority have strayed.

In a democracy, however limited, whoever believes that a given law or regulation is unjust has the right, nay the moral and civic duty, to oppose it by peaceful means such as complaining or petitioning,

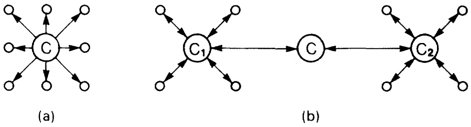


Fig. 11.3. Authoritarian centralization: central domination. Democratic centralization: central coordination of decentralized units.

litigating or taking part in rallies. Under a political dictatorship the fight against injustice is illegal and therefore either indirect, through front organizations, or direct, through open confrontation. The latter can run from civil disobedience *à la* Thoreau or Gandhi to riot or revolution.

In theory social revolutions are quick and effective surgical means that spare us much misery. In practice the cost of most social revolutions — in lives taken or shattered, and in ideals distorted or corrupted — has been staggering. It is not true that, once power has been seized, everything else follows easily and quickly. Nor is it true that any social revolution is bound to accomplish more than any social reform: many a social reform has achieved much that certain revolutions have yet to deliver. Besides, the systematic recourse to violence is not precisely a school for either democracy or morals: these are usually the first casualties of sustained violence. Still, revolution, if it has a chance of succeeding, is morally justified if and only if it is the only means of toppling an utterly unjust and inflexible government.

Defense and war come next in our agenda. Hardly anybody disputes the norm that one of the main specific functions of the state is to provide national security. The only point in dispute is how best to achieve this goal: whether by buying arms or making friends, by strengthening defense proper or preparing for aggression — or even launching a "preventive" war to forestall foreign aggression.

Since only enemies make war, it stands to reason that the best way of preventing war is to turn enmity into friendship through intense exchanges of goods, people and ideas, the way France and Germany have been doing since the end of WWII. Partners may compete in some regards but only by peaceful means. Hence partnership is the best insurance against war. On the other hand weapons, particularly of the offensive kind, buy insecurity. And the more offensive weapons a government accumulates, the more likely it is to attack or to be attacked out of greed or fanaticism, stupidity or plain computer error. Never underestimate the obtuseness of professional warriors or the probability of machine failure. And never count on the moral conscience of either.

The current arms race is foolish and immoral. It is ruining humankind at the rate of $1.8 million a minute. It will burden our posterity, if any, with a huge fiscal deficit; it is depleting the fossil fuel reserves and polluting the environment; and it contributes to unemployment, because the arms industry employs only few and highly qualified workers. Moreover the arms race is distorting all values by making national security, rather than welfare, the overriding concern of the state. The obsession with national security provides an excuse for militarism, nationalism, state terrorism, the repression of dissent, the erosion of social services, and even education for war. It obfuscates judgment and blunts moral sensibility.

The nuclear arms race is the most monstrous of all races. It is foolish because the existing nuclear stockpile suffices to kill every human being fifty times over, and because most of the nuclear warriors themselves do not seem to believe that it will ever be used. Still, *the* (not just *a*) nuclear war might break out at any moment because of fanaticism or by mistake. And this permanent threat poses the greatest moral problem ever faced by humankind, for — as the Rev. Theodore M. Hesburgh (1983) has said — if nuclear war were to break out there would be no human beings left and therefore no moral problems to be tackled. The only practically and morally acceptable solution to the nuclear arms race is total multilateral nuclear disarmament. Anything short of this is suicidal sophistry. In particular, nuclear deterrence is risky, ruinous, and immoral.

To sum up, it is a mistake to believe that he who buys weapons buys security. Security cannot be bought unilaterally, but it may be earned by mutual agreement. It can be earned by (*a*) making honest efforts to resolve conflicts through negotiation, (*b*) universal disarmament, starting with nuclear, chemical and biological weapons, (*c*) intense trade and cultural exchange and cooperation, (*d*) education aimed at forming peaceful citizens of the world rather than brutal jingoists, and (*e*) a hundred-fold strengthening of the UN to the point of turning it into the world government. If you want war, prepare for it; if you want peace, make it.

A federation of all nations — the United States of the World —

would have to be managed (not dominated) by a world government. The main specific function of this government of governments would be to address global problems, such as the preservation of peace through disarmament and negotiation, the protection of the biosphere, and the rational exploitation of the non-renewable resources. It should also engage in megaprojects requiring the cooperation of many governments, such as the reclamation of deserts, the relocation of refugees and excess population, fusion research, the mining of ocean floors and of the moon, space exploration, the development of appropriate or medium level technologies, and the facilitation of the flow of goods, people and information throughout the world, as well as the support of regional development plans.

The formation of a world federal government would gradually erase the international borders — which is what the European Economic Community is doing. Where there are no frontiers it become pointless to defend them by force of arms. Where every government is a partner in a world-wide enterprise, the obsession with national security is replaced with enthusiasm for international exchange and cooperation. Nations do not disappear within the USW but nationalism, a major source of international conflict, is replaced with cosmopolitanism. The concept of humankind, so far only a biological one, becomes a biosociological notion. The Stoic and Cynic *cosmopolis* comes finally within reach.

It goes without saying that membership in the USW would impose certain restrictions on national sovereignty, such as the banning of offensive weapons, a limitation on mining, and the enforcement of waste management regulations. But this would be nothing new, as every government relinquishes some of its sovereignty the moment it becomes a member of an international body or signs an international treaty. It is one thing to surrender national sovereignty by opening the door to a conqueror or to a transnational company intent on plundering natural resources, and quite a different matter to relinquish some sovereignty in exchange for security and cooperation, progress and, above all, survival.

To conclude, it is high time for political philosophy to become subordinated to moral philosophy, and politics to morals. This is not only a matter of righteousness but also of prudence, for the divergence between morals and politics, and between moral and political philosophy, has led humans to the brink of extinction. (Remember Franklin's dictum: If scoundrels knew the advantages of virtue, they would be virtuous out of sheer roguery.) This divergence would not exist in a just society managed by a technodemocratic state, for in it (*a*) every citizen would participate in politics, so that there would be no political class; (*b*) the elected officers would hold office only for limited periods and they would consider themselves, as Jefferson (1807) told A. v. Humboldt, as public property; (*c*) there would be intense cooperation and consultation instead of strong leadership — which is the preface to tyranny; and (*d*) the public servants, whether elected or appointed, would be sensitive to criticisms and suggestions by private citizens, nongovernmental organizations, and the press, so that they would not dare hindering social development but would, on the contrary, be ready to facilitate it.

5. CULTURAL

Every society, however primitive, has a culture, i.e. a subsystem composed of people who produce or help produce, diffuse or consume cultural items — scientific, technological, humanistic, artistic, or ideological. See Figure 11.4. (For an analysis of culture as a concrete system see Vol. 4, Ch. 5, Sects. 2.4 and 3.2.)

The level of development of any culture depends to some extent on that of the associated economy and polity. Without some economic prosperity and some political freedom there can be no vigorous — i.e. widespread and quickly changing — modern culture. Indeed, the latter involves certain nonutilitarian activities, such as basic scientific and humanistic research, that require some material means and some freedom of innovation and communication. It was no accident that capitalism begat modern science, science-based technology, and political democracy. Nor is it an accident that such business paradises as Hong Kong, Singapore, Taiwan and Seoul have yet to give us the likes of Cervantes, Beethoven, Einstein, and Russell.

The market economy has an ambivalent effect on culture. By turning certain cultural items into marketable commodities, it stimulates their production and socialization. But by the same token it favors the mass production of low grade and sometimes adulterated cultural items, while it neglects basic science, the humanities and genuine art. Potboilers for mass consumption may sell better than refined masterpieces; and short-term and narrow R&D that promises quick returns gets more support than any long-term adventurous, hence uncertain, exploration of the unknown. In the process the deepest sources of creation — curiosity, the joy and fun of trying something new, and the

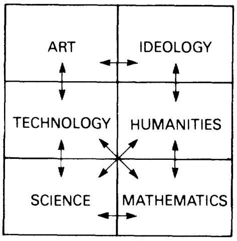


Fig. 11.4. The contemporary cultural system and its subsystems. The arrows represent both information flows and controls. Redrawn from Figure 5.8 in Vol. 4.

passion for beauty — dry up. As a consequence the river of culture dwindles and becomes muddy. In culture as elsewhere the Market giveth and the Market taketh away.

Because business cannot help being profit-oriented, it cannot take care of the whole of culture. Hence the state and non-governmental cultural associations must design and carry out a cultural policy involving the protection of the seemingly superfluous (i.e. economically unprofitable) cultural activities — precisely those that distinguish civilization from barbarism.

Every cultural (or countercultural) policy presupposes some philosophy of culture, i.e. some system of ideas concerning the nature, function and value of culture. In turn, a philosophy of culture has ontological, epistemological, axiological and ethical underpinnings. In particular, if the ethical root of a philosophy of culture happens to be some egoistic ethical doctrine, the corresponding cultural policy will be elitist. That is, it will proclaim that culture is only for those who can afford to pay for it, whence higher education, museums, libraries, and the like should be privatized or left to the good will of wealthy patrons. And wherever greed is the dominant force, philistinism prevails.

In a modern political democracy the bulk of higher culture is managed by the state. In an integral democracy it would be managed by the culture workers themselves organized in cooperatives and in autonomous state corporations such as the BBC. Only such nongovernmental and state-supported autonomous organizations can promote economically unprofitable cultural activities and heterodoxy. On the other hand business companies and authoritarian (or simply meddlesome) governments tend to reward only short-term practical pursuits and orthodoxy, or even intolerance. Where philistinism and conformism prevail, culture withers.

Every creative individual is heterodox in some respects. Hence heterodoxy should be encouraged: justified criticism and valuable innovation should be rewarded. On the other hand we should reject the global rejection of modern culture and the attempts to replace it with myth, superstition, or pop culture. In particular we should reject bogus science (e.g. parapsychology), medical quackery (e.g. homeopathy), pseudophilosophy (e.g. existentialism), and bogus art (e.g. tachisme). It is not that we need a thought police, or a Ministry of Truth and Beauty, for either would suffocate creativity and thus stifle culture. But we do need scholars, teachers and journalists capable of diagnosing junk culture and alerting the public to it, as well as laws criminalizing medical and psychological quackery for being hazardous to health and pocket.

The state and the nongovernmental cultural organizations have the duty to protect cultural creativity, i.e. to support innovative scientists, humanists and artists — and even some technologists, for industry is often slow in recognizing the potential of technological innovation, and even more so that of scientific innovation. The state has this duty because it must protect all basic human rights, and the rights to create and to learn are among them. Actually, in a rapidly changing world like ours, learning and innovating are duties as well as rights, because participation and mutual help — which constitute the mortar of society — are inefficient or even counterproductive without a modicum of expertise and creativity. Therefore learning ought to be regarded as a lifelong process, and access to learning as a basic right at any age. Consequently massive and intensive continuing education is in the interest of everyone except political tyrannosaurs.

In a modern society the state provides opportunities for university education. In particular, it provides higher education in the basic sciences and humanities, which are of little use to business companies. The state must continue to do so because it can be the most powerful custodian of culture, and there is no modern culture without science and the humanities. On the other hand it is doubtful that the state has the obligation to foot the bill of the training of engineers, physicians, lawyers, accountants and other professionals who are likely to engage in profitable practice upon graduating. It would be unfair to force the taxpayer to contribute to their education. But it would also be unfair to restrict admission to professional schools to the wealthy. The fair thing to do is to offer future professionals loans repayable out of their professional earnings. On the other hand good arts and sciences students should get free education.

Scientists, humanists and artists deserve state support not only for their original contributions, which enrich our lives, but also because they spread culture as teachers, authors, journalists, or performers. In particular, some of them are needed to train engineers, physicians, lawyers, and business administrators. Hence if a politician should fail to understand that the nonutilitarian branches of culture have an intrinsic value, in that they enhance the quality of our lives, we may try to make him understand that they are needed in any good school.

Education should be seen in its various functions: as a means for self-realization, as part of the socialization process, as part of the formation of responsible and cooperative individuals, as the training of skilled workers, and as preparation to face the unexpected challenges posed by social evolution — or to make them happen. Of all the imaginable social orders technodemocracy is the one that calls for the most demanding education, because every normal member of such society is supposed to take part in some ongoing decision making process or other. If every decision had to be taken by a committee of inexperts, rather than by a team of knowledgeable individuals advised by experts, people would spend most of their lives attending meetings rather than working and enjoying themselves.

What holds for schools holds also, *mutatis mutandis*, for mass media. These too discharge several functions: they broadcast information, make value, opinion and taste, promote prosocial (or antisocial) behavior, awake (or smother) vocation, and stimulate (or inhibit) consumption. In order to operate efficiently the media need freedom: if subjected to censorship they only output propaganda. However, it is no secret that the lack of economic freedom can have the same effect. In particular, when the news is handled as one more commodity it can be sold (hence socialized), bought (hence withheld), or adulterated. In an integral democray the media would be owned and operated by co-operatives, and they would serve a more enlightened public because they would enlighten rather than drug it.

The socialization of higher culture that began five centuries ago is still midway. In fact, in most societies higher culture is the monopoly of an elite. The masses are usually served the cultural equivalent of junk food: pulp magazines and commercial television, instant art, pseudoscience, and even pseudophilosophy. In an integral democracy there would be no such division of culture into genuine and fake, one for elites and the other for the masses. In such a society everyone would have the opportunity to obtain a good scientific, technological, humanistic and artistic education, so that few if any people would produce or consume cultural junk. True, since only a minority would avail themselves of this opportunity, the culture would be dominated by elites. But these would be cultural, not economic or political elites.

6. SUMMARY

The survival of humankind is being threatened by a handful of manmade global problems: nuclear war and the subsequent nuclear winter (or at least fall), the depletion of natural resources, environmental degradation, poverty, the North-South inequality, and overpopulation. Since global problems call for global solutions, they may only be solved by the concerted action of all national governments. (See e.g. Brown & Postel 1987 p. 18 and WCED 1987.)

The solution to the above-mentioned global problems requires more than a few international agreements and emergency measures. It calls for world-wide, comprehensive and radical social reforms if only because we have been pushed to the brink of extinction by allowing economic and political greed to trample on the most basic values and norms, starting with the right to well-being and interpersonal and international solidarity. We need a new social order centered in the moral norm "Enjoy life and help live". This social order is integral technodemocracy, for it combines social justice with technical efficiency.

Integral technodemocracy is superior to its main rivals, the relief or liberal state, and statist or bureaucratic socialism. The former was introduced to save capitalism from its own greed by palliating its outstanding faults, namely unemployment and the inequitable distribution of wealth. (See e.g. Keynes 1936.) The relief state has alleviated much misery but it has not eradicated the main sources of misery, in particular war and the preparation for war. Moreover it has now reached the end of its tether in nations with high unemployment rates and heavy military expenditures. As for statist or bureaucratic socialism, it has most of the defects of the liberal megastate and few of its virtues. In particular it has disowned most civil rights, concentrated political power, and restrained creativity in the name of a largely obsolete ideology.

Any big government, whether capitalist or socialist, is bound to shrink the private sphere, smother initiative and creativity, and erode self-reliance and solidarity, while favoring "the development of unbridled, narrow-minded individualism" (Kropotkin 1902 p. 227). However, the solution to the problem of big government does not consist in miniaturizing, let alone in eliminating it altogether. The rational thing to do is to change some of the functions of the state, to decentralize and democratize it, and to work for a just society and a peaceful international order.

This goal cannot be attained overnight because there are no "quick fixes" for social problems. There are none because all important social issues have deep roots in the past and in all social spheres — the economy, the culture and the polity. Social problems call for long-term social reforms. However, these cannot be of the piecemeal kind because (*a*) every society is a multidimensional system and (*b*) partial reforms do not question the basic social order. We need global reforms in the two senses of the term 'global': embracing all of the subsystems of every society, and world-wide. As the historian Carr (1961 p. 207) wrote, "Progress in human affairs [. . .] has come mainly through the bold readiness of human beings not to confine themselves to piecemeal improvements in the way things are done, but to present fundamental changes in the name of reason to the current way of doing things and to the avowed or hidden assumptions on which it rests". And all great social changes have been fueled by moral passions combined with material interests (Moore 1973 p. 3).

Great changes for the better cannot be improvised: they must be planned. Free rational individuals have life plans, and independent rational governments have national projects. A national project may be tacit or explicit, and it may be pushed by special interests groups or by the entire society. The success or failure of a national project depends on the ability of a group of political and cultural leaders to persuade their fellow citizens that they should devote their best energies to implementing the plan. "National projects only succeed when they can meet the needs of the people, are compatible with the vital interests of other people, and take the social, economic, technological, moral and cultural forces of the times into account" (Stern 1967 p. 270). Nowadays we need not only feasible national projects, particularly in the underdeveloped nations. We also need a planetary project capable of saving humankind from itself, and of building a viable, peaceful and just social order. In other words, we need a world federal government heading — not muscling — the United States of the World.

To conclude. The pessimist, the conformist, and the indifferent may claim that this entire chapter is an exercise in political fiction. We rejoin that, though futuristic, our sketch for a viable world is feasible and that, unless humankind embarks on some such project, it will have no future. Those who refuse to see that we are on the brink of extinction, or who do realize this but do not bother to do anything, condone the greatest evil that human beings could ever imagine.

CHAPTER 12

VALUES AND MORALS FOR A VIABLE FUTURE

If H. G. Wells were alive today he might write as follows, though of course in better style. "As Charles Dickens might say, ours is the worst of times and also the best of times. It is the worst of times because humankind is quickly digging its own grave. It is the best of times because never before has humankind had so many powerful means to bring about a bright future. The lemmings jump off the cliff to their deaths because they have never heard of continental drift. But we know better, so we need not keep on marching blindly. We are still in time to turn away from the cliff. But the time is getting shorter by the day."

In this chapter we shall review the current global crisis that is threatening the survival of humankind. We shall suggest that humans have only themselves to blame for having chosen the wrong values and therefore the wrong morals. And we shall end by suggesting a system of values and moral norms that may save us from ourselves. In the process we shall sketch the kind of society that would incorporate such values and morals.

1. THE THIRTEEN HORSEMEN OF THE APOCALYPSE

Anyone who watches the daily newscast and has a nodding acquaintance with such statistical yearbooks as the Worldwatch Institute's *State of the World* and Sivard's *World Military and Social Expenditures* realizes that humankind is on the brink of an apocalyptic catastrophe. This time the horsemen of the Apocalypse are for real and they are thirteen, and they are galloping fast and laying our world to waste. They are the following (Bunge 1988b):

(1) *the arms race*, which may end in the final holocaust and nuclear winter, or may go on with conventional weapons even after the entire nuclear stockpile has been dismantled, and in the meantime is frightening and impoverishing us all;

(2) *the rapid depletion of nonrenewable resources*, which include oil (the reserves of which may last another two decades or so) and all the critical industrial metals;

(3) *environmental degradation* as a result of industrial, urban and domestic fumes and toxic wastes, and of the destruction of forests;

(4) *shrinking and misuse of arable land* due to deforestation, overcultivation, abuse of chemical fertilizers, herbicides and defoliants, and the cultivation of plants for cattle fodder and alcoholic beverages;

(5) *overindustrialization* in the rich countries, due to overconsumption, and *underindustrialization* and underconsumption in the poor ones;

(6) *overpopulation*: the population of most nations keeps growing beyond their carrying capacities and social resources, sometimes doubling every generation;

(7) *hunger and thirst*: about 800 million people are not eating enough for an active working life, and twice as many drink contaminated water;

(8) *unemployment and underemployment*: only three out of five adults in the Third World are fully employed;

(9) *inadequate health care*: most people in the world have no access to modern sanitation and medicine;

(10) *militarization*: the statesmen of most nations worry more about national security than about peace and welfare, and more than half of the so-called developing nations are ruled by the military;

(11) *debt*: most nations have accumulated huge national and foreign debts, usually contracted to pay for military expenditures, which will have to be repaid by posterity, and in the meantime increase poverty and dependency;

(12) *the widening North-South gap* as a result of the mismanagement of the economy and the state in the poor countries, and of the ruthless exploitation of their resources on the part of some industrialized nations;

(13) *the decline of culture*: an increasing number of well-trained brains are devoted to weapons design or to the maintenance of inefficient and oppressive governments, while basic science, the humanities and the arts, as well as the civilian technologies, are declining while mass-produced junk culture is prospering and displacing genuine folk culture.

All these problems are moral as well as social. They touch on all the aspects of contemporary life and they affect, directly or indirectly, all nations regardless of their level of development. They are global problems because the biosphere does not honor any national borders and because, from 1492 on, all nations have become interdependent. If any country suffers deforestation, desertification, massive pollution, or plague, its neighbors are affected as well. If any country is ravaged by poverty, foreign debt, civil war, or international conflict, its neighbors suffer as well. Just as all organisms are components of a single food web, so all societies are part of a world wide network, and every person has five billion partners. In short, the world is physically, biologically and socially one regardless of its many divisions. Hence the predicament of any nation is the predicament of all.

2. OLD AND NEW VALUES AND MORALS

We all live by our values. If our values are right we may be saved, otherwise we shall succumb. Now, I submit that the prevailing value systems are wrong, and that this is the root of our woes.

The prevailing value systems are axiological individualism (egocentrism) and holism (sociocentrism). The former supports a morality that places private interests before the public interest, whereas holism sacrifices the individual to the whole. According to individualism what is good for me (or for my business company) is good enough for all. According to holism what is good for the whole (or the state or the cause) is good for anyone.

In either case wants are given precedence over basic needs and legitimate aspirations. So, we reach for what we (or our rulers or employers) desire, rather than for what we need to survive, let alone to enjoy well-being in good conscience. We do not always desire what is objectively valuable but, on the contrary, we often value what we (or our rulers or employers) desire. In either case we lose sight of the Socratic ideal of *the good individual in a good society.*

The consequences of such value systems for morality are obvious. In one case we are told to do as we please, in the other that we may only do what "society" (meaning the state, church or party) wants. In the first case we are told that we are fully autonomous, in the second that we are fully heteronomous. Accordingly in the first case we are expected to behave as spoiled brats, in the second as kamikazes. Fortunately normal people are neither.

Individualism fosters consumerism, which is depleting nonrenewable resources, fouling our nest, and exploiting the Third World. Holism breeds passivity and smothers creativity and initiative. And both generate the obsession with national security, a major cause of militarism and waste. The egocentric sacrifices the future for the sake of the present, and the sociocentric does the converse, and between the two they have pushed humankind to the brink.

If our diagnosis is correct, i.e. if the survival of humankind is at risk because some of us — mainly our leaders and habit makers — have espoused the wrong values and consequently the wrong morals, then the way to salvation is clear. We must stop identifying the good with whatever is likely to benefit an individual at the expense of his fellow human beings, or society at the expense of its individual components.

If we realize that the future of our species is at stake, and wish it to be saved, we must transvaluate the dominant value systems. We must reject all of the obsolete ideologies and return to the basic values: to whatever may help us meet our basic needs and legitimate wants — *ours* not just *mine* or *theirs.* The corresponding moral code can be summarized into our Norm 4.6: *Enjoy life and help live.* Indeed, only by satisfying our basic needs and legitimate (i.e. socially admissible) aspirations can we enjoy life; and only by helping others pursue the same goal may we get the help we need and keep a clean conscience.

3. A SURVIVAL AND DEVELOPMENT MORALITY

The *summum bonum* of the value system advocated in this book is human survival. Since at present the survival of humankind is threatened by the nuclear arms race and the degradation of the environment, stopping these two degenerative processes is the supreme moral imperative of our time: See Figure 12.1.

Everything else must be subordinated to this imperative for, unless it is observed, there will be nobody left to pursue any human goals, not even evil ones. In particular, the socialism-capitalism (or East-West) conflict is not the main one. The main alternative today is survival or extinction.

If the preceding is true, then the greatest evildoers of our time are not so much the criminals within the reach of the criminal codes, as the warmongers and the large scale plunderers and polluters, as well as their political accomplices. By comparison the terrorists and murderers, swindlers and thieves, and even the mafiosi and drug pushers are small fry for, after all, they are not out to steal the planet or exterminate all life. The power of those large scale criminals must be drastically curtailed if humankind is to survive.

However, surviving is not enough. All human beings and, indeed, all

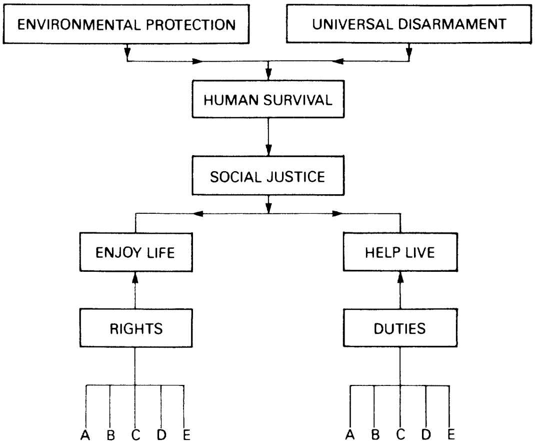


Fig. 12.1. Basic schema of agathonism. A = Environmental, B = Biological, C = Cultural, D = Political, E = Economic. An arrow represents the converse of the implication relation. Thus environmental protection and universal disarmament are jointly necessary for the survival of humankind, while the exercise of human rights is necessary for the enjoyment of life. The three upper tiers summarize social morality, the three lower ones personal morality.

animals strive for a good life. Moreover, human beings must be able at least to imagine what the good life is like, if they are to bother doing something to save humankind. We need then more than a salvific morality: we need a more comprehensive moral code encouraging us to seek the good for ourselves and others — i.e. an *agathonist* morality. (Ἀγαθόν = good.) And what is good? Anything capable of meeting a need or a want (Ch. 1, Sect. 3). Whence our supreme Norm 4.6: *Enjoy life and help live.*

Since what is at stake is the whole of humankind and, indeed, the entire biosphere, we had better look for a universal or cross-cultural morality rather than for a sectarian one. A universal moral code would have to be a *minimal* or *core* morality, i.e. it would have to equal the intersection of all the existing viable moralities. Recall Figure 7.2. Beyond such minimal or core morality every human group should be free to adopt its own moral code, preferably the one best adjusted to its lifestyle or occupation.

It is not hard to construct a set of moral norms that would receive the assent of all human beings with the sole exception of pathological egotists — a minority in every viable society. In fact, most people prefer pleasure to pain, sufficiency to scarcity, company to solitude, peace to war, security to insecurity, employment to idleness, cooperation to strife, freedom to serfdom, fairness to injustice, benevolence to malevolence, loyalty to betrayal, knowledge to ignorance, truth to falsity, and so on. And all people can learn to put up with some burdens in exchange for the benefits they receive: i.e. they can learn to do their duty provided they are not required to perform useless sacrifices.

The characteristics we have just listed are part of human nature (Ch. 5, Sect. 2.1). They are cross-cultural or universal, and they are the roots of a universal morality, i.e. one that could be binding on all human beings. Table 12.1 exhibits the skeleton of a possible *Universal Charter of Human Rights and Duties* constituting the core morality we need to survive and develop. To be sure no morality suffices for anything, but some morality is necessary in any society. Without moral principles we drift, and with bad ones we sink; only when equipped with good moral principles can we hope to reach the other shore.

4. A VIABLE SOCIAL ORDER

The morality sketched in Table 12.1 can only be observed in a good society, i.e. one where everybody can get what she needs, and is free to pursue her legitimate interests and inclinations. None of the existing societies is good enough, for in every one of them some group or other is being discriminated against biologically, economically, culturally, or politically. (Obvious examples: Unemployment and underemployment, economic barriers to health care and higher education, political persecution and apathy.) A good society is just, i.e. free from biological, economic, cultural or political privileges. (The converse is false: a just society may be miserable, hence not good: Recall Definition 2.6.) In a just society, or orthopolis, everyone has the same opportunities, although not everyone may be able or willing to make full use of them all.

Moreover a just society is integrally democratic, i.e. everyone has an effective say in the management of all the social groups she is a member of. Unless integral, a democracy is unjust and unstable. For example, political democracy is compatible with race and gender discrimination,

TABLE 12.1. Sample of human rights and duties. Alphabetic ordering

Right to	Duty to
ask for help	offer help
associate	refrain from ganging up for mischief
choose friends	be loyal to friends
choose honest lifestyle	respect other people's lifestyles
choose honest occupation	help others find occupation
choose spouse	discharge marital obligations
compete	cooperate
create	respect valuable creations
criticize errors & wrongs	help correct errors & right wrongs
defend one's rights	respect other people's rights
elect officers	help duly elected officers do their job
enjoy the environment	protect the environment
equality before the law	observe the just laws
equality of opportunity	allow others to seize their opportunities
equal pay for equal work	deserve one's wages
free inquiry	be intellectually honest
free speech	refrain from inciting antisocial behavior
health care	contribute to the health care system
learn	share knowledge
live	respect other living beings
live with parents	share in domestic chores or expenses
love	respect and protect the beloved
meet one's basic needs & legitimate desires	help others meet their basic needs & legitimate desires
own things for personal use	take good care of one's property
participate	participate
plan one's own life	help others fulfil their life projects
privacy	respect other people's privacy
pursue one's interests	help other people pursue their interests
old age security	contribute to the pension fund
rest and leisure	respect other people's rest and leisure
run for office	do a good job in office
security	help maintain security
travel	help fellow travelers
work	observe the rules of good workmanship
worship	tolerate other people's religious beliefs and practices

economic exploitation, and cultural marginality. Partial democracy, whether political, economic, or cultural, is elitist and therefore rife with social conflicts — hence repressive, unstable, or both. The four components of integral democracy hang together, i.e. they form a system: See Figure 12.2.

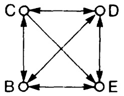


Fig. 12.2. The four components of democracy hang together. B = Biological democracy, E = Economic democracy, C = Cultural democracy, D = Political democracy.

However, no matter how complete, unless accompanied by expertise democracy is inefficient, hence primitive or wasteful. And nowadays expertise is not just craftsmanship, in particular political skill, but science-based technology, in particular sociotechnology — for which see Vol. 7, Ch. 5, Sect. 4. Therefore the social ideal to work for in the light of the above is a combination of integral democracy with technology, i.e. what we have called *integral technodemocracy.*

An integral technodemocracy will not stand still: it will advance because its members will be free and competent to tackle new problems as well as to improve on old solutions. However, given the current population excess and the huge gap between the haves and the have-nots (individuals and nations), we should not count on an economy of plenty for all. Except in the realm of culture, where benefits always exceed costs, we must expect temperance rather than abundance — unless we continue to be intent on disinheriting our posterity.

Given the increasing interdependence among nations, and the fact that all of them are afflicted with both global and local problems, it is desirable to work for a world federation headed by a world government capable of tackling planetary problems (Ch. 11, Sect. 4). Such political unity need not eliminate regional diversities. On the contrary, it should help preserve every valuable diversity by sparing people the ravages of war, famine and plague, and by helping the victims of natural disaster — which are likely to increase in frequency and severity, at least for a while, as a result of pollution, deforestation, overcultivation, and other concomitants of unbridled reproduction, production and consumption.

Moreover, a world government, to be effective and lasting, would have to be democratic, for otherwise it would not obtain the required cooperation of the member states. In particular it would have to tolerate differences in economic and political organization, and it would have to guarantee each member nation the right to opt out of any world program — except of course for the offensive arms control and the global environmental protection programs.

The old social orders are incompatible with our supreme moral norm *Enjoy life and help live* and, a fortiori, with the Universal Charter of Human Rights and Duties sketched in Table 12.1. Moreover none of those orders promises a viable future. The alternatives are not capitalism or socialism, and national sovereignty or superpower hegemony, but integral democracy or elitism, and world anarchy or international coordination and cooperation.

5. CONCLUSION

The morality advocated in this book is based on a value theory according to which anything that promotes welfare is good. Our morality can be summed up in the norm *Enjoy life and help live*, which is selftuist rather than either purely selfish or purely altruistic. The morality in question has, among others, the following characteristics. It is

(i) *agathonist*: it recommends the pursuit of the good for self and others, rather than the exclusive pursuit of pleasure, profit or power — or the grim discharging of heavy and thankless duty;

(ii) *systemist* rather than either individualist (egocentric) or holist (sociocentric);

(iii) *consequentialist*, though not utilitarian, for maintaining that our actions are to be judged mainly by their consequences;

(iv) *empiricist*, or sensitive to the tests of real life, rather than aprioristic or conventionalist;

(v) *rationalist*, or open to argument, rather than dogmatic (in particular intuitionist);

(vi) *scientistic* for holding that science and science-based technology are the best judges of claims to good and right, as well as the best antidotes against axiological and ethical dogmatism;

(vii) *realistic* for tackling real life problems and proposing practical norms;

(viii) *universalist* rather than nationalist, parochial or sectarian;

(ix) *democratic* for holding that the human condition can be improved only through active public participation in the enjoyment and management of all the natural, economic, cultural and political resources of society;

(x) *secular* or humanistic rather than supernaturalist or revealed.

I submit that, far from being a futile academic exercise, a pious wish, or a cloak of narrow vested interests, a morality of this sort is necessary for the survival and development of the human species. To philosophize we must be alive, but if humankind is to survive, and a fortiori to develop, we must philosophize correctly.

This concludes the *Treatise*. May its readers feel motivated to improve on it.

BIBLIOGRAPHY

Ackerman, B. (1980). *Social Justice in the Liberal State.* New Haven, CT: Yale University Press.

Ackoff, R. L. and F. E. Emery (1972). *On Purposeful Systems.* Chicago: Aldine.

Afriat, S. N. (1987). *Logic of Choice and Economic Theory.* Oxford: Clarendon Press.

Agassi, J. (1977). *Towards a Rational Anthropology.* The Hague: Martinus Nijhoff.

Agassi, J. (1985). *Technology. Philosophical and Social Aspects.* Dordrecht and Boston: Reidel.

Aigner, T. and R. L. Balster (1978). Choice behavior in Rhesus monkeys: cocaine versus food. *Science* **201**: 534—535.

Albert, H. (1979). *Das Elend der Theologie.* Hamburg: Hoffmann und Campe.

Albert, H. (1985). *Treatise on Critical Reason.* Princeton, NJ: Princeton University Press.

Alchourrón, C. E. and E. Bulygin (1971). *Normative Systems.* Wien-New York: Springer-Verlag.

Alexander, R. D. (1987). *The Biology of Moral Systems.* New York: Aldine De Gruyter.

Alexander, S. (1933). *Beauty and Other Forms of Value.* London: Macmillan.

Allais, M. and O. Hagen, Eds. (1979). *Expected Utility Hypotheses and the Allais Paradox.* Dordrecht-Boston: Reidel.

Almeder, R. (1983). Morality in the marketplace. In Snoeyenbos, Almeder and Humber, Eds., pp. 84—93.

Alwin, D. F. (1987). Distributive justice and satisfaction with material well-being. *American Sociological Review* **52**: 83—95.

Anand, P. (1987). Are the preference axioms really rational? *Theory and Decision* **23**: 189—214.

Aristotle. *The Nichomachean Ethics.* Transl. D. Ross. Oxford-New York: Oxford University Press.

Aristotle. *Politics.* Transl. B. Jowett. New York: Modern Library.

Aronfred, J. (1980). *Conduct and Consciousness.* New York: Academic Press.

Arrow, K. (1951). *Social Choice and Individual Values.* New York: John Wiley.

Axelrod, R. (1984). *The Evolution of Cooperation.* New York: Basic Books.

Axelrod, R. and W. D. Hamilton (1981). The evolution of cooperation. *Science* **211**: 1390—1396.

Ayer, J. A. (1936). *Language, Truth &Logic.* Oxford: Oxford University Press.

Baier, A. (1986). Trust and antitrust. *Ethics* **96**: 231—260.

Baier, K. (1966). *The Moral Point of View: A Rational Basis of Ethics.* New York: Random House.

Bandura, A. (1977). *Social Learning Theory.* Englewood Cliffs, NJ: Prentice-Hall Inc.

Barbour, I. G. (1980). *Technology, Environment, and Human Values.* New York: Praeger.

Batson, C. D. and J. S. Coke (1983). Empathic motivation of helping behavior. In J. T. Cacioppo and R. E. Petty, Eds., *Social Psychophysiology: A Sourcebook*, pp. 417—433. New York and London: Guilford Press.

Beer, S. (1966). *Decision and Control.* New York: Wiley.

Bellon, B. and J. Niosi (1988). *The Decline of the American Economy.* Montreal-New York: The Black Rose.

Bentham, J. (1789). *An Introduction to the Principles of Morals and Legislation.* J. H. Burns and H. L. A. Hart, Eds., London and New York: Methuen, 1982.

Bergson, A. (1967). Market socialism revisited. *J. Political Economy* **75**: 655—673.

Bergson, H. (1932). *Les deux sources de la morale et de la religion.* Paris: Presses Universitaires de France.

Bidney, D. (1953). The concept of value in modern anthropology. In L. Kroeber, Ed., *Anthropology Today.* Chicago: University of Chicago Press.

Black, M. (1964). The gap between "is" and "should". *Philosophical Review* **73**: 165—181.

Blanshard, B. (1961). *Reason and Goodness.* London: Allen & Unwin.

Blasi, A. (1980). Bridging moral cognition and moral action, *Psychological Bulletin* **88**: 1—45.

Bodde, D. L. (1986). Ethics and policy analysis. *Technology in Society* **8**: 149—155.

Bonner, J. (1986). *Politics, Economics and Welfare.* Brighton: Harvester Press.

Boochkin, M. (1982). *The Ecology of Freedom.* Palo Alto, CA: Cheshire Books.

Borovoy, A. A. (1988). *When Freedoms Collide.* Toronto: University of Toronto Press.

Bouglé, C. (1922). *Leçons de sociologie sur l'évolution des valeurs.* Paris: A. Colin.

Brandt, R. B. (1959). *Ethical Theory.* Englewood Cliffs NJ: Prentice-Hall.

Brandt, R. B. (1979). *A Theory of the Good and the Right.* Oxford: Clarendon Press.

Brecht A. (1959). *Political Theory.* Princeton NJ: Princeton University Press.

Broad, C. D. (1930). *Five Types of Ethical Theory.* London: Routledge & Kegan Paul.

Bronowski, J. (1959). *Science and Human Values.* New York: Harper.

Brown, L. R. and S. Postel (1987). Thresholds of change. In Brown *et al.* 1987 pp. 3—19.

Brown, L. R. *et al.* (1987). *State of the World 1987. A Worldwatch Institute Report on Progress Towards a Sustainable Society.* New York: W. W. Norton and Co.

Brown, L. R. *et al.* (1989). *State of the World 1989. A Worldwatch Institute Report on Progress Toward a Sustainable Society.* New York: W. W. Norton and Co.

Bruera, J. J. (1945). La lógica, el derecho y la Escuela de Viena. *Minerva* **2**: 170—177.

Buber Agassi, J. (1980). Toward a rational work ethics. *Crossroads, International Dynamics, and Social Change* **6**: 75—108.

Buber Agassi, J. (1986). Dignity in the workplace: Can work be dealienated? *Journal of Business Ethics* **5**: 271—284.

Buchanan, A. (1985). *Ethics, Efficiency, and the Market.* Totowa NJ: Rowman & Allanheld.

Buchanan, J. M. (1954). Social choice, democracy, and free market. *American Economic Review* **44**: 114—123.

Buchanan, J. M. (1975a). *The Limits of Liberty: Between Anarchy and Leviathan.* Chicago: University of Chicago Press.

Buchanan, J. M. (1975b). The Samaritan's dilemma. In Phelps, Ed., pp. 71—85.

Buchanan, J. M. (1979). *What Should Economists Do?* Indianapolis: Liberty Press.

Buchanan, J. M. and R. D. Tollison Eds. (1972). *The Theory of Public Choice: Political Applications of Economics*. Ann Arbor, MI: University of Michigan Press.

Buchanan, J. M. and G. Tullock (1962). *The Calculus of Consent*. Ann Arbor: University of Michigan Press.

Bunge, A. (1915). *El culto de la vida*. Buenos Aires: Juan Perrotti.

Bunge, C. O. (1919). *Estudios filosóficos*. Buenos Aires: La Cultura Argentina.

Bunge, M. (1960). *Etica y ciencia*. Buenos Aires: Siglo Veinte.

Bunge, M. (1961). Ethics as a science. *Philosophy and Phenomenological Research* **20**: 139—152.

Bunge, M. (1962a). An analysis of value. *Mathematicae Notae* **18**: 95—107.

Bunge, M. (1962b). *Intuition and Science*. Englewood Cliffs, NJ: Prentice-Hall. Repr.: Westport, CT: Greenwood Press, 1975.

Bunge, M. (1973a). A decision theoretic model of the American War in Vietnam. *Theory and Decision* **3**: 323—338.

Bunge, M. (1973b). *Method, Model and Matter*. Dordrecht: Reidel.

Bunge, M. (1975). Towards a technoethics. *Philosophic Exchange* **2**(2): 69—79. Repr.: *The Monist* **60**: 96—107 (1977).

Bunge, M. (1976). A model for processes combining competition with cooperation. *Applied Mathematical Modeling* **1**: 21—23.

Bunge, M. (1980). *The Mind-Body Problem*. Oxford-New York: Pergamon Press.

Bunge, M. (1981). *Scientific Materialism*. Dordrecht-Boston: D. Reidel Publ. Co.

Bunge, M. (1982a). *Economía y filosofía*. Madrid: Tecnos.

Bunge, M. (1982b). The revival of causality. In G. Fløistad, Ed., *Contemporary Philosophy* Vol. 2 pp. 133—155. The Hague: Martinus Nijhoff.

Bunge, M. (1988a). Basic science is innocent, applied science and technology can be guilty. In G. A. Lemarchand & A. R. Pedace Eds. *International Symposium on Scientists, Peace, and Disarmament*, pp. 245—261. Singapore: World Scientific.

Bunge, M. (1988b). The thirteen riders of the Apocalypse. *Free Inquiry* **8**(3) 9.

Bunge, M. (1989). *Development and the Environment*. In E. F. Byrne and J. Pitt, Eds., *Technological Transformation* pp. 285—304. Dordrecht: Kluwer.

Butler, J. (1726). *Sermons on Human Nature*. In *The works of J.B.*, vol. 2. Oxford: Clarendon Press, 1874.

Caja Laboral Popular (1984). El conjunto cooperativo de Mondragón. *Revista Interamericana de Autogestión y Acción Comunal* **3**: 189—222.

Cajal, S. Ramón y (1941). *Los tónicos de la voluntad*. Madrid: Espasa-Calpe.

Calvocoressi, P. (1988). *A Time for Peace*. London: Hutchinson.

Camus, A. (1951). *L'homme révolté*. Paris: Gallimard.

Carens, J. H. (1981). *Equality, Moral Incentives, and the Market*. Chicago: University of Chicago Press.

Carr, E. H. (1961). *What is History*? London: Macmillan. Repr. Penguin 1964.

Chagnon, N. A. (1988). Life histories, blood revenge, and warfare in a tribal population. *Science* **239**: 985—992.

Chazan, B., Ed. (1985). *Contemporary Approaches To Moral Education*. New York: Teachers College Press.

Constant, B. (1797). *Des réactions politiques*. In Écrits et discours politiques, 2 v. O. Pozzo di Borgo, Ed. Paris: Jean-Jacques Pauverts, 1964.

Cragg, W., Ed, (1983). *Contemporary Moral Issues.* Toronto-New York: McGraw-Hill Ryerson.

Dahl, R. A. (1985). *A Preface to Economic Democracy.* Berkeley and Los Angeles: University of California Press.

Darwin, C. (1871). *The Descent of Man.* New York: Collier, 1902.

Dawkins, R. (1976). *The Selfish Gene.* Oxford: Oxford University Press.

Dickens, C. (1861). *Great Expectations.* Harmondsworth: Penguin Books, 1965.

Dujovne, L. (1959). *Teoría de los valores y filosofía de la historia.* Buenos Aires: Paidos.

Durkheim, E. (1895). *De la division du travail social.* 6th. ed.: Paris: PUF, 1960.

Durkheim, E. (1924). *Sociologie et philosophie.* Paris: Alcan.

Dworkin, R. (1978). *Taking Rights Seriously.* Cambridge MA: Harvard University Press.

Dworkin, G. (1979). Paternalism. In Laslett and Fishkin Eds. pp. 78—96.

Ebling, F. J. G., Ed. (1969). *Biology and Ethics.* New York: Academic Press.

Eccles, J. C. and D. N. Robinson (1985). *The Wonder of Being Human.* Boston and London: New Science Library (Shambhala).

Eichner, A. S., Ed. (1983). *Why Economics is not yet a Science.* New York: Sharpe, Armonk.

Einstein, A. (1950). The laws of science and the laws of ethics. Foreword to P. Frank, *Relativity: A Richer Truth,* pp. v—viii. Boston: Beacon Press.

Elden, J. M. (1981). Political efficacy at work. *Amer. Political Science Review* **75**: 43—58.

Elster, J. and A. Hylland, Eds. (1966). *Foundations of Social Choice Theory.* Cambridge: Cambridge University Press; Olso: Universitetsforlaget.

Engels, F. (1878). *Anti-Dühring.* Moscow: Foreign Languages Publishing House.

Estrin, D. and J. Swaigen (1978). *Environment on Trial.* Toronto: Macmillan.

Evarts, E. V., Y. Shinoda and S. P. Wise (1984). *Neurophysiological Approaches to Higher Brain Function.* New York: Wiley (Interscience).

Feinberg, J. (1984). *The Moral Limits of the Criminal Law, Vol. 1, Harm to Others.* New York: Oxford University Press.

Feinberg, J. (1985). *The Moral Limits of the Criminal Law,* Vol. 2, *Offense to Others.* New York: Oxford University Press.

Feinberg, J. (1986). *The Moral Limits of the Criminal Law,* Vol. 3, *Harm to Self.* New York: Oxford University Press.

Ferrater-Mora, J. and P. Cohn (1981). *Etica aplicada. Del aborto a la violencia.* Madrid: Alianza Editorial.

Fishkin, J. S. (1984). *Beyond Subjective Morality: Ethical Reasoning and Political Philosophy.* New Haven: CT: Yale University Press.

Forman, F. (1989). *The Metaphysics of Liberty.* Dordrecht-Boston: Kluwer.

Forsyth, D. R. and R. E. Berger (1982). The effects of ethical ideology on moral behavior. *Journal of Social Psychology* **117**: 53—56.

Fox, A. (1974). *Beyond Contract: Work, Power and Trust Relations.* London:

Freedman, B. (1978). A meta-ethics for professional morality. *Ethics* **89**: 1—19.

Freud, S. (1913). *Totem and Taboo.* In *Works,* J. Strachey and A. Freud, Eds., London: Hogarth Press, 1955.

Freud, S. (1923). *The Ego and the Id.* In *Works*, J. Strachey and A. Freud, Eds., London: Hogarth Press 1962.

Friedman, M. (1962). *Capitalism and Freedom.* Chicago & London: University of Chicago Press.

Friedman, M. (1970). The social responsibility of business is to increase its profits. *The New York Times Magazine* 1970.09.13 pp. 33, 122—126.

Frondizi, R. (1963). *What is Value*? La Salle, ILL: Open Court.

Frondizi, R. (1977). *Introducción a los problemas fundamentales del hombre.* México: Fondo de Cultura Económica.

Gale, S. (1974). On a metatheory of planning theory. Continuing Series on Planning Theory, No. 1, Dept. of City and Regional Planning, University of Pennsylvania.

Garzón Valdés, E. (1986a). Moral y política. *Cuadernos de ética* **1**: 7—31.

Garzón Valdés, E. (1986b). Los deberes positivos generales y su fundamentación. *Doxa* **3**: 17—33.

Garzón Valdés, E. (1987). Es éticamente justificable el paternalismo jurídico? *Revista Latinoamericana de Filosofía* **13**: 279—298.

Gasparski, W. and W. Pszczolowski, Eds. (1983). *Praxiological Studies: Polish Contributions to the Science of Efficient Action.* Dordrecht: Reidel.

Gauthier, D. (1982). No need for morality: the case of the competitive market. *Philosophic Exchange* **3**: 41—54.

Gauthier, D. (1984). Deterrence, maximization, and rationality. *Ethics* **94**: 474—495.

Gauthier, D. (1986). *Morals by Agreement.* Oxford: Clarendon Press.

Geach, P. T. (1956). Good and evil. *Analysis* **17**: 33—42.

Gewirth, A. (1978). *Reason and Morality.* Chicago: University of Chicago Press.

Ghai, D. P., A. R. Khan, E. L. H. Lee and T. Alfthan (1977). *The Basic Needs Approach to Development: Some Issues Regarding Concepts and Methodology.* Geneva: International Labour Office.

Gibbard, A. (1984). Health care and the prospective Pareto principle. *Ethics* **94**: 261—282.

Gilbert, J. O., R. J. Light and F. Mosteller (1975). Assessing social innovations: an empirical base for policies. In C. A. Bennett and A. A. Lumsdaine, Eds., *Evaluation and Experiment: Some Critical Issues in Assessing Social Programs* pp. 39—193. New York: Academic Press.

Gilligan, C. (1982). *In a Different Voice: Psychological Theory and Women's Development.* Cambridge MA: Harvard University Press.

Gilligan, J. (1976). Beyond morality: Psychoanalytic reflections on shame, guilt and love. In T. Lickona, Ed., *Moral Development and Behavior*, New York: Holt, Rinehart & Winston.

Glover, J. (1986). "It makes no difference whether or not I do it". In Singer, Ed., pp. 125—144.

Goslan, D. A., Ed. (1969). *Handbook of Socialization Theory and Research.* Chicago: Rand McNally.

Gouldner, A. J. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review* **25**: 161—178.

Graham, L. R. (1981). *Between Science and Values.* New York: Columbia University Press.

Griffin, J. (1982). Modern utilitarianism. *Revue Internationale de Philosophie* **36**: 331—375.

Grousec, J. E. and T. Dix (1986). The socialization of prosocial behavior: theory and reality. In Zahn-Waxler, Cummings and Iannotti, Eds., pp. 218—237.

Guyau, J. M. (1885). *Esquisse d'une morale sans obligation ni sanction.* 2nd ed. Paris: F. Alcan, 1890.

Hahn, F. (1982). On some difficulties of the utilitarian economist. In Sen and Williams, Eds., pp. 187—198.

Halal, W. E. (1986). *The New Capitalism.* New York: John Wiley & Sons.

Hamilton, W. D. (1964). The genetical theory of social behaviour. *J. Theoret. Biol.* **7**: 1—52.

Hampshire, S. (1983). *Morality and Conflict.* Cambridge, MA: Harvard University Press.

Handy, R. (1969). *Value Theory and the Behavioral Sciences.* Springfield ILL: Charles C Thomas.

Hardin, R., J. J. Mearsheimer, G. Dworkin, and R. E. Godin, Eds. (1985). *Symposium on Ethics and Nuclear Deterrence, Ethics* **95**: 409–739.

Hare, R. M. (1952). *The Language of Morals.* London: Oxford University Press.

Hare, R. M. (1963). *Freedom and Reason.* Oxford: Oxford University Press.

Harris, J. (1975). The survival lottery. *Philosophy* **50**: 81—87.

Harris, M. (1979). *Cultural Materialism: The Struggle for a Science of Culture.* New York: Random House.

Harman, G. (1977). *The Nature of Morality.* New York: Oxford University Press.

Harsanyi, J. C. (1958). Ethics in terms of hypothetical imperatives. *Mind* **67**: 305—316.

Harsanyi, J. C. (1976). *Essays on Ethics, Social Behavior, and Scientific Explanation.* Dordrecht: Reidel.

Harsanyi, J. C. (1982). Morality and the theory of rational behavior. In A. K. Sen and B. Williams, Eds., *Utilitarianism and Beyond* pp. 42—44. Cambridge: Cambridge University Press.

Harsanyi, J. C. (1983). Basic moral decisions and alternative concepts of nationality. *Social Theory and Practice* **9**: 231—244.

Harsanyi, J. C. (1985). Does reason tell us what moral code to follow and, indeed, to follow any moral code at all? *Ethics* **96**: 42—55.

Hart, H. L. A. (1961). *The Concept of Law.* Oxford: Oxford University Press.

Hart, H. L. A. and A. M. Honoré (1959). *Causation in the Law.* Oxford: Clarendon Press.

Hartmann, (1949). *Ethik.* Berlin: de Gruyter.

Harvey, P. H. and K. Talls (1986). Do animals avoid incest? *Nature* **320**: 575—576.

Hayek, F. A. (1944). *The Road to Serfdom.* London: Routledge & Kegan Paul.

Hayek, F. A. (1976). *Law, Legislation and Liberty,* Vol. 2: *The Mirage of Social Justice.* Chicago: University of Chicago Press.

Hebb, D. O. (1953). On motivation and thought. In *Contributions à l'étude des sciences de l'homme* **2**: 41—47. Repr. in H. A. Buchtel, Ed., *The Conceptual Nervous System.* Oxford: Pergamon, 1982.

Hebb, D. O. (1974). What psychology is about. *American Psychologist* **29**: 71—79.

Hegel, G. W. F. (1837). *Vorlesungen über die Philosophie der Geschichte.* In *Werke,* v. 12. Frankfurt: Suhrkamp, 1970.

Herbst, P. G. (1974). *Socio-Technical Design.* London: Tavistock.

Hesburgh, T. M. (1983). Address to the Association of Universities and Colleges of Canada and the American Council on Education, Toronto, October 13, 1983.

Hobbes, T. (1651). *Leviathan.* London: Dent, 1953.

Hoffman, M. W. and J. Miller Moore, Eds. (1984). *Business Ethics: Readings and Cases in Corporate Morality.* New York: McGraw-Hill.

Homans, G. C. (1967). *The Nature of Social Science.* New York: Harcourt, Brace & World.

Hospers, J. (1982). *Human Conduct: Problems of Ethics.* New York: Harcourt Brace Jovanovich Inc.

Hospers, J. (1988). Ecology and liberty. *Liberty* Sept. 1988: 23—31.

Hörz, H. E. and U. Wilke, Eds. (1986). *Ethik.* Berlin: VEB Deutscher Verlag der Wissenschaften.

Hull, C. L. (1944). Value, valuation, and natural-science methodology. *Phil. Sci.* **11**: 125—141.

Hume, D. (1751). *An Enquiry Concerning the Principles of Morals.* Indianapolis: Hackett, 1983.

Huxley, T. H. (1893). Evolution and ethics. In Huxley and Huxley pp. 60—102.

Huxley, T. H. and J. Huxley (1947). *Evolution and Ethics.* London: Pilot Press.

Ingenieros, J. (1917). *Hacia una moral sin dogmas.* Buenos Aires: L. J. Rosso y Cia.

International Labour Office (1976). *Employment, Growth, and Basic Needs: A One-World Problem.* Geneva: ILO.

Jantsch, E., Ed. (1969). *Perspectives in Planning.* Paris: OECD.

Kamin, L. J. (1974). *The Science and Politics of IQ.* New York: Wiley.

Kamin, L. J. (1986). Is crime in the genes? *Scientific American* **254** (**2**): 22—27.

Kanger, S. (1957). New foundations for ethical theory. In R. Hilpinen, Ed., *Deontic Logic: Introductory and Systematic Reading*, pp. 36—57. Dordrecht: Reidel, 1970.

Kant, I. (1785). *Grounding for the Metaphysics of Morals.* Indianapolis: Hackett, 1981.

Kant, I. (1792). Reflexionen zur Religionsphilosophie. In *Gesammelte Schriften* v. XIX. Berlin: Walter de Gruyter, 1934.

Kant, I. (1794). Letter to Friedrich Wilhelm II. In *Briefwechsel von Imm. Kant*, Vol. 3, pp. 50—53. Munich: Georg Müller, 1913.

Kaufmann, W. (1973). *Without Guilt and Justice.* New York: Wyden.

Kautsky, K. (1906). *Ethics and the Materialist Conception of History*, 4th ed. Transl. J. B. Askew. Chicago: Charles H. Kerr & Co., s.d.

Kelsen, H. (1949). *General Theory of Law and State.* Cambridge, MA: Harvard University Press.

Keynes, J. M. (1936). *The General Theory of Employment, Interest and Money.* In *Collected Writings*, Vol. VII. London: Macmillan & Cambridge University Press, 1973.

Kitcher, P. (1985). *Vaulting Ambition: Sociobiology and the Quest for Human Nature.* Cambridge MA: MIT Press.

Knight, F. (1935). *The Ethics of Competition.* London: George Allen & Unwin.

Knoke, D. (1988). Incentives in collective action organizations. *Amer. Sociol. Rev.* **53**: 311—329.

Kohlberg, L. (1981). *Essays on Moral Development.* Vol. I: *The Philosophy of Moral Development.* New York: Harper & Row.

Kohlberg, L. (1984). *Essays on Moral Development.* Vol. II: *The Psychology of Moral Development.* New York: Harper & Row.

Kolnai, A. (1938). *The War against the West.* London: Gollancz; New York: Viking Press.

Kolnai, A. (1978). *Ethics, Value, and Reality.* Indianapolis: Hackett.

Korsgaard, C. M. (1986). Aristotle and Kant on the source of value. *Ethics* **96**: 486—505.

Koshland, D. D., Jr. (1977). A response regulator model in a simple sensory system. *Science* **196**: 1055—1063.

Koslowski, P. (1986). *Ethik des Kapitalismus,* 3rd ed. Tübingen: J.C.B. Mohr.

Kotarbinski, T. (1965). *Praxiology: An Introduction to the Sciences of Efficient Action.* Oxford: Pergamon.

Kraft, V. (1937). *Die Grundlagen einer wissenschaftlichen Wertlehre.* Vienna: Julius Springer. Transl.: *Foundations for a Scientific Analysis of Value.* Dordrecht: Reidel, 1981.

Kranzberg, M., Ed. (1980). *Ethics in an Age of Pervasive Technology.* Boulder, CO: Westview Press.

Kropotkin, P. (1902). *Mutual Aid.* Boston: Extending Horizon Books, 1955.

Kropotkin, P. (1924). *Ethics: Origin and Development.* London: George G. Harrap & Co.

Kurtz, P. (1973a). What is humanism? In Kurtz, Ed., pp. 1—14.

Kurtz, P. (1973b). The individual, the organization, and participatory democracy. In Kurtz, Ed., pp. 189—209.

Kurtz, P., Ed. (1973). *Moral Problems in Contemporary Society. Essays in Humanistic Ethics.* Buffalo NY: Prometheus Books.

Kurtz, P. (1988). *Forbidden Fruit.* Buffalo, NY: Prometheus Books.

Küppers, B.-O. (1982). Der Verlust aller Werte. *Natur* **4**: 65—74.

Ladd, J. (1970). Morality and the ideal of rationality in formal organizations. *The Monist* **54**: 488—516.

Lamont,W. D. (1981). *Law and the Moral Order.* Aberdeen: Aberdeen University Press.

Landsell, H. (1988). Laboratory animals need only humane treatment: Animal "rights" may debase human rights. *Intern. J. Neuroscience* **42**: 169—178.

Lane, R. E. (1981). Markets and politics: The human product. *British Journal of Political Science* **11**: 1—16.

Lee, V. (1912). *Vital Lies: Studies of Some Varieties of Recent Obscurantism,* 2 vols. London: John Lane.

Lemmon, E. J. (1965). Deontic logic and the logic of imperatives. *Logique et analyse* **8**: 39—71.

Lenin, V. I. (1920). The tasks of the Young Leagues. In *Selected Works* Vol. II pp. 661—674. Moscow: Foreign Languages Publishing House, 1947.

Leonard, H. S. (1959). Interrogatives, imperatives, truth, falsity, and lies. *Philosophy of Science* **26**: 172—186.

Lepper, M. R., D. Greene and R. Nisbett (1973). Undermining children's intrinsic interest with extrinsic reward: A test of the overjustification hypothesis. *Journal of Personality and Social Psychology* **28**: 129—137.

Lepper, M. R., G. Sagotsky, J. Dafoe and D. Greene (1982). Consequences of superfluous social constraints: Effects on young children's social inferences and subsequent intrinsic interest. *Journal of Personality and Social Psychology* **42**: 51—65.

Lévy-Bruhl, L. (1903). *La morale et la science des moeurs*. Paris: Alcan, 1937.

Lewis, C. I. (1946). *An Analysis of Knowledge and Valuation*. La Salle, ILL: The Open Court Publ. Co.

Lewis, C. I. (1955). *The Ground and Nature of the Right*. New York: Columbia University Press.

Lewis, C. I. (1969). *Values and Imperatives: Studies in Ethics*. J. Lange, Ed., Stanford: Stanford University Press.

Lewis, C. S. (1943). *The Abolition of Man*. Glasgow: Collins: 1978.

Lewis, W. A. (1955). *The Theory of Economic Growth*. London: George Allen & Unwin.

Lewontin, R. C., S. Rose and L. Kamin (1984). *Not in our Genes*. New York: Pantheon.

Locke, J. (1690). *Two Treatises of Government*. P. Laslett, Ed., Cambridge: Cambridge University Press, 1970.

Lorenz, K. (1970). *Studies in Animal and Human Behaviour*, 2 vols. Cambridge MA: Harvard University Press.

Luce, R. D. and H. Raiffa (1957). *Games and Decisions*. New York: Wiley & Sons.

Lumsden, J. C. and E. O. Wilson (1981). *Genes, Mind and Culture*. Cambridge MA: Harvard University Press.

Lyttelton, A., Ed. (1973). *Italian Fascisms: From Pareto to Gentile*. London: Jonathan Cape.

Machina, M. J. (1987). Decision-making in the presence of risk. *Science* **236**: 537—543.

MacIntyre, A. (1984). *After Virtue*, 2nd ed. Notre Dame, IN: University of Notre Dame Press.

Mackie, J. L. (1977). *Ethics. Inventing Right and Wrong*. Hardmondsworth: Penguin Books.

Malinowski, B. (1944). *A Scientific Theory of Culture and other Essays*. Chapel Hill: University of North Carolina Press.

March, J. G. and H. A. Simon (1958). *Organizations*. New York: John Wiley.

Margolis, H. (1984). *Selfishness, Altruism, and Rationality: A Theory of Social Choice*. Chicago and London: University of Chicago Press.

Marković, M. (1982). *Democratic Socialism: Theory and Practice*. New York: St. Martin's Press.

Marcović, M. (1987). Radical democracy: A rational reconstruction. *Grazer Philosophische Studien* **30**: 121—138.

Martínez-Selva, J. M. (1981). Aspecto psicológico. In J. M. Vázquez, Ed., *Los gitanos en Murcia, hoy, 1980*. Murcia: Universidad de Murcia.

Maslow, A. (1968). *Towards a Psychology of Being*, 2nd ed. Princeton NJ: Van Nostrand.

Max-Planck-Gesellschaft (1984). *Verantwortung und Ethik in der Wissenschaft*. München: Max-Planck-Gesellschaft.

May, K. O. (1954). Intransitivity, utility, and the aggregation of preference patterns. *Econometrica* **22**: 1—13.

May, R. M. (1988). How many species are there on Earth? *Science* **241**: 1441—1449.

Maynard Smith, J., Ed. (1982). *Evolution Now*. London: Macmillan; San Francisco: Freeman.

McHale, J. and M. McHale (1977). *Basic Human Needs: A Framework for Action.* Houston, TX: University of Houston.

Mencken, H. L. (1934). *Treatise on Right and Wrong.* New York: Alfred A. Knopf.

Mencken, H. L. (1956). *Minority Report.* New York: Alfred A. Knopf.

Michalos, A. (1980). *North American Social Report* 5 vols. Dordrecht-Boston: Reidel.

Mill, J. S. (1852). *Principles of Political Economy,* 3rd ed. In *Collected Works* Vols. II and III. Toronto: University of Toronto Press, 1965.

Mill, J. S. (1859). *On Liberty.* In *Collected Works* Vol. XVIII. Toronto: University of Toronto Press, 1977.

Mill, J. S. (1863). *Utilitarianism.* In *Collected Works,* Vol. X. Toronto: University of Toronto Press, 1969.

Mill, J. S. (1873). *Autobiography.* J. J. Coss, Ed., New York: Columbia University Press, 1924.

Miller, D. (1976). *Social Justice.* Oxford: Clarendon Press.

Miller, D. (1977). Socialism and the market. *Political Theory* **5**: 473—489.

Miller, D. (1980). Jerusalem not yet built: A reply to Lesnoff on capitalism, socialism and democracy. *Political Studies* **28**: 584—589.

Miró Quesada, F. (1951). La lógica del deber ser y su eliminabilidad. *In Ensayos de filosofía del derecho,* pp. 114—124. Lima: Universidad de Lima, 1986.

Moessinger, P. (1975). Developmental study of fair division and property. *European J. Social Psych.* **5**: 385—394.

Moessinger, P. (1987). Let's be as precise as possible: An interview with Mario Bunge. *New Ideas in Psychol.* **5**: 385—397.

Moessinger, P. (1989). *La psychologie morale.* Paris: Presses Universitaires de France.

Monod, J. (1971). *Chance and Necessity.* New York: Knopf.

Montesquieu, Ch.-L. de S. (1748). *De l'esprit des lois.* In *Oeuvres complètes,* Paris: Masson, 1950—55.

Moore, B., Jr (1973). *Reflections on the Causes of Human Misery and Upon Certain Proposals to Eliminate them.* Boston: Beacon Press.

Moore Lappé, F. and J. Collins (1988). *World Hunger: 12 Myths.* London: Earthscan Publications.

Moore, G. E. (1903). *Principia Ethica.* Cambridge: Cambridge University Press.

Mosteller, F. (1981). Innovation and evaluation. *Science* **211**: 881—886.

Mosterín, J. (1978). *Racionalidad y acción humana.* Madrid: Alianza Editorial.

Muller, E. N. (1988). Democracy, economic development, and income inequality. *American Sociological Review* **53**: 50—68.

Murdock, G. P. (1949). *Social Structure.* New York: Macmillan.

Mussolini, B. and G. Gentile (1932). The doctrine of fascism. In A. Lyttelton, Ed., *Italian Fascisms from Pareto to Gentile,* pp. 39—57. London: Jonathan Cape, 1973.

Myers, D. B. (1986). The principle of reversibility: Some problems of interpretation. *J. Value Inquiry* **20**: 19—28.

Myrdal, G. (1942). *An American Dilemma.* Repr.: New York, Harper & Row, 1962.

Myrdal, G. (1969). *Objectivity in Social Research.* New York: Pantheon Books.

Nader, R., R. Petkas, and K. Blackwell, Eds. (1972). *Whistle Blowing.* New York: Grossman.

Nagel, E. (1968). The enforcement of morals. *The Humanist* **28**: 20—27.

Nagel, T. (1979). *Mortal Questions.* Cambridge: Cambridge University Press.

Neurath, O. (1945). Review of Hayek's *The Road to Serfdom* 1944. In *Gesammelte* *philosophische und methodologische Schriften*, Vol. 2, pp. 979—981. Wien: Hölder-Pichler-Tempsky.

Neyrinck, J. (1986). *Le huitième jour de la création*. Lausanne: Presses Polytechniques Pomandes.

Nielsen, K. (1984). *Equality and Liberty: A Defense of Radical Egalitarianism*. Totowa NJ: Rowman & Allanheld.

Nienhaus, V. (1984). *Kontroversen um Markt und Plan*. Darmstadt: Wissenschaftliche Buchgesellschaft.

Nietzsche, F. (1886). *Beyond Good and Evil*. In *Basic Writings of N.*, W. Kaufmann, Ed. New York: Modern Library, 1968.

Nietzsche, F. (1888). *The Will to Power*. W. Kaufmann, Ed. New York: Vintage Books, 1968.

Nietzsche, F. (1892). *Thus Spake Zarathustra*. London: Dent, 1958.

Nino, C. S. (1985). *Etica y derechos humanos*. Buenos Aires: Paidos.

Nozick, R. (1974). *Anarchy, State and Utopia*. New York: Basic Books.

Nriagu, J. O. and J. M. Pacyna (1988). Quantitative assessment of worldwide contamination of air, water and soils by trace metals. *Nature* **333**: 134—139.

Olds, J. and P. Milner (1954). Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain. *J. Comp. Physiol. Psychol.* **47**: 419—427.

Olson, M. (1982). *The Rise and Decline of Nations*. New Haven, CN: Yale University Press.

Olweus, D., J. Block and M. Radke-Yarrow, Eds. (1986). *The Development of Antisocial and Prosocial Behavior*. New York: Academic Press.

Opp, K.-D. (1982). The evolutionary emergence of norms. *Brit. J. Social Psych.* **21**: 139—149.

Ossowska, M. (1970). *Social Determinants of Moral Ideas*. Philadelphia: University of Pennsylvania Press.

Otis, L. S. and J. A. Cerf (1963). Conditioned avoidance learning in two fish species. *Psychological Reports* **12**: 679—682.

Panksepp, J. (1986). The psychobiology of prosocial behaviors: separation distress, play, and altruism. In Zahn-Waxler, Cummings & Iannotti, Eds., pp. 19—57.

Parfit, D. (1984). *Reasons and Persons*. Oxford: Oxford University Press.

Parra-Luna, F. (1983). *Elementos para una teoría formal del sistema social*. Madrid: Editorial de la Universidad Complutense.

Pascal, L. (1986). Judgement Day. In Singer, Ed., pp. 105—123.

Perry, R. B. (1926). *General Theory of Value*. Cambridge, MA: Harvard University Press, 1954.

Phelps, E. S., Ed. (1975). *Altruism, Morality, and Economic Theory*. New York: Russell Sage Foundation.

Piaget, J. (1932). *The Moral Judgment of the Child*. London: Kegan Paul, Trench, Trubner.

Plato. *The Dialogues of Plato*, 2 vols. B. Jowett, transl. New York: Random House, 1937.

Pollock, J. L. (1986). A theory of moral reasoning. *Ethics* **96**: 506—523.

Pollock, L. (1988). Evaluating moral theories. *American Philosophical Quarterly* **25**: 229—240.

Popper, K. R. (1966). *The Open Society and its Enemies*, 5th ed. London: Routledge & Kegan Paul.

Potter, N. T. and M. Timmons, Eds. (1985). *Morality and Universality.* Dordrecht-Boston: Reidel.

Pugh, G. E. (1976). *The Biological Origin of Human Values.* N.Y.: Basic Books.

Quinton, A. (1989). *Utilitarian Ethics*, 2nd ed. London: Duckworth.

Rand, A. (1964). *The Virtue of Selfishness: A New Concept of Egoism.* New York: American Library.

Rapoport, A. (1983). *Mathematical Models in the Social and Behavioral Sciences.* New York: Wiley.

Ratner, L. G. (1984). The utilitarian imperative: Autonomy, Reciprocity, and Evolution. *Hofstra Law Review* **12**: 123—137.

Rawls, J. (1971). *A Theory of Justice.* Cambridge MA: Belknap Press.

Rawls, J. (1985). Justice as fairness: Political, not metaphysical. *Philosophy and Public Affairs* **14**: 223—251.

Regan, D. (1980). *Utilitarianism and Co-operation.* Oxford: Clarendon Press.

Reichenbach, H. (1959). *Modern Philosophy of Science. Selected Essays.* London: Routledge & Kegan Paul.

Rescher, N. (1966). *Distributive Justice.* Indianapolis: Bobbs-Merrill.

Rescher, N. (1969). *Introduction to Value Theory.* Englewood Cliffs NJ: Prentice-Hall.

Rescher, N. (1975). *Unselfishness.* Pittsburgh: University of Pittsburgh Press.

Rescher, N. (1987). *Forbidden Knowledge.* Dordrecht and Boston: Reidel.

Robinson, J. (1962). *Economic Philosophy.* Harmondsworth: Penguin, 1964.

Robinson, J. and J. Eatwell (1973). *An Introduction to Modern Economics*, rev. ed. London: McGraw Hill.

Rogers, C. (1964). Toward a modern approach to values: the valuing process in the mature person. *Journal of Abnormal and Social Psychology,* **68**: 160—167.

Rohatyn, F. G. (1987). The blight on Wall Street. *New York Review of Books* March 12, 1987, pp. 21—23.

Ross, W. D. (1930). *The Right and the Good.* Oxford: Clarendon Press.

Rousseau, J.-J. (1755). Discours sur l'origine et les fondements de l'inégalité parmi les hommes. In *The Social Contract and Discourses*, G. D. H. Cole transl. London: Dent, 1973.

Rushton, J. P. (1980). *Altruism, Socialization, and Society.* Englewood Cliffs, NJ: Prentice-Hall.

Russell, B. (1949). *Authority and the Individual.* New York: Simon and Schuster.

Russell, B. (1954). *Human Society in Ethics and Politics.* London: George Allen & Unwin.

Russell, E. S. (1976). Report of the *ad hoc* committee [on genetics, race, and intelligence]. *Genetics* **83**: s99—s101.

Ruyer, R. (1952). *Philosophie de la valeur.* Paris: Armand Colin.

Sandler-Girbau, H. R. (1977). *Alquileres e inflación.* México: Editorial Jus.

Scanlon, T. M. (1982). Contractualism and utilitarianism. In Sen & Williams Eds. pp. 103—128.

Schaub, J. H. and K. Pavlovic, Eds. (1983). *Engineering Professionalism and Ethics.* New York: John Wiley & Sons.

Scheffler, S., Ed. (1988). *Consequentialism and its Critics.* Oxford: Oxford University Press.

Schelling, T. (1960). *The Strategy of Conflict.* Cambridge, MA: Harvard University Press.

Schelling, T. (1968). Game theory and the study of ethical systems. *J. Conflict Resolution* **12**: 33—44.

Schlick, M. (1925). *Allgemeine Erkenntnislehre*, 2nd ed. Engl. transl. A. E. Blumberg, *General Theory of Knowledge.* New York-Wien: Springer-Verlag, 1971.

Schlick, M. (1930). *Fragen der Ethik.* Wien: Julius Springer. Engl. transl.: *The Problems of Ethics.* New York: Prentice-Hall, 1938.

Schmidt, P. F. (1959). Ethical norms in scientific method. *Journal of Philosophy* **56**: 644—652.

Schrecker, T. F. (1984). *Political Economy of Environmental Hazards.* Ottawa: Law Reform Commission of Canada.

Schweikart, D. (1980). *Capitalism or Worker Control*? New York: Praeger.

Scitovsky, T. (1976). *The Joyless Economy.* Oxford: Oxford University Press.

Seebass, G. and R. Tuomela, Eds. (1985). *Social Action.* Dordrecht and Boston: Reidel.

Selucky, R. (1979). *Marxism, Socialism, Freedom: Towards a General Democratic Theory of Labour-Management Systems.* New York: St Martin's Press.

Sen, A. K. (1970). *Collective Choice and Social Welfare.* San Francisco: Holden-Day.

Sen, A. K. (1977). Rational fools: A critique of the behavioral foundations of economic theory. *Philosophy and Public Affairs* **6**: 317—344.

Sen, A. (1979). Utilitarianism and welfarism. *Journal of Philosophy* **76**: 463—489.

Sen, A. (1982). *Poverty and Famines: An Essay on Entitlement and Deprivation.* Oxford: Clarendon Press.

Sen, A. (1985). Well-being, agency and freedom. *Journal of Philosophy* **82**: 169—221.

Sen, A. (1987). *On Ethics and Economics.* Oxford: Basil Blackwell.

Sen, A. and B. Williams, Eds. (1982). *Utilitarianism and Beyond.* Cambridge: Cambridge University Press; Paris: Ed. de la Maison des Sciences de l'Homme.

Settle, T. (1976). *In Search of a Third Way.* Toronto: McClelland & Stewart.

Shirer, W. L. (1960). *The Rise and Fall of the Third Reich.* New York: Simon and Schuster.

Shrader-Frechette, K. (1981). *Environmental Ethics.* Pacific Grove, CA: Boxwood Press.

Sidgwick, H. (1907). *The Methods of Ethics* 7th ed. Preface by J. Rawls. Indiana: Hackett, 1982.

Simpson, G. G. (1964). *This View of Life.* New York: Harcourt, Brace and World.

Singer, P. (1975). *Animal Liberation.* New York: A New York Review Book.

Singer, P., Ed. (1986). *Applied Ethics.* Oxford: Oxford University Press.

Sivard, R. L. (1987). *World Military and Social Expenditures 1987—88*, 12th ed. Washington: World Priorities.

Sjoberg, G., Ed. (1967). *Ethics, Politics, and Social Research.* Cambridge MA: Schenkman Publ. Co.

Skinner, B. F. (1971). *Beyond Freedom and Dignity.* New York: Alfred A. Knopf.

Smart, J. J. C. (1973). *An Outline of a System of Utilitarian Ethics.* In Smart and Williams, pp. 1—74.

Smart, J. J. C. and B. Williams (1973). *Utilitarianism for and against.* Cambridge: Cambridge University Press.

Smart, R. N. (1958). Negative utilitarianism. *Mind* NS **67**: 542—543.

Sjoberg, G., Ed., (1967). *Ethics, Politics and Social Research.* Cambridge, MA: Schenkman.

Smith, A. (1759). *Theory of Moral Sentiments.* New York: A. M. Kelley, 1966.

Smith, A. (1776). *An Inquiry Into the Nature and Causes of the Wealth of Nations.* Chicago: University of Chicago Press, 1976.

Snow, C. P. (1959). *The Two Cultures and the Scientific Revolution.* Cambridge: Cambridge University Press.

Snoyenbos, M., R. Almeder, and J. Humber, Eds. (1983). *Business Ethics. Corporate Values and Society.* Buffalo, N.Y.: Prometheus Books.

Spicker, S. F. and H. T. Engelhardt Jr, Eds. (1977). *Philosophical Medical Ethics: Its Nature and Significance.* Dordrecht-Boston: Reidel.

Spinoza, B. (1670). *Theologico-Political Treatise.* In *Chief Works of Spinoza.* New York: Dover, 1951.

Spinoza, B. (1677). *Ethics.* In *Chief Works of Spinoza.* New York: Dover, 1951.

Spurgeon, D., Ed. (1979). *Give us the Tools.* Ottawa: International Development Research Centre.

Stent, G. S., Ed. (1980). *Morality as a Biological Phenomenon,* rev. ed. Berkeley: University of California Press.

Stern, A. (1936). *La philosophe des valeurs: Regards sur ses tendences actuelles en Allemagne.* Paris: Hermann.

Stern, A. (1967). *Geschichtsphilosophie und Wertproblem.* München: Ernst Reinhardt.

Stevenson, C. L. (1946). *Ethics and Language.* New Haven CN: Yale University Press.

Stevenson, R. L. (1886). The strange case of Dr Jekyll and Mr Hyde. In *The Strange Case of Dr Hekyll and Mr Hyde and Other Stories.* Harmondsworth: Penguin, 1979.

Stojanović, S. (1988). *Perestroika: From Marxism and Bolshevism to Gorbachev.* Buffalo, NY: Prometheus Books.

Stone, C. (1974). *Should Trees have Standing?* Los Altos, CA: William Kaufmann.

Storer, M. B., Ed. (1980). *Humanist Ethics.* Buffalo: Prometheus Books.

Streeten, P. and S. Burki (1978). Basic needs: Some issues. *World Development* **6**: 411—421.

Suárez, F. (1612). *Tractatus de legibus ac Deo legislatore.* In *Selections from Three Works by Francisco Suárez, S.J.* Oxford: Oxford University Press, 1944.

Taylor, J. F. A. (1958). The masks of society: The grounds of obligation in the scientific enterprise. *The Journal of Philosophy* **55**: 485—502.

Taylor, M. (1976). *Anarchy and Cooperation.* New York: Wiley.

Thatcher, R. W., R. A. Walker and S. Giudice (1987). Human cerebral hemispheres develop at different rates and ages. *Science* **236**: 1110—1113.

Thibodeau, F. R. and H. H. Field, Eds. (1984). *Sustaining Tomorrow: A Strategy for World Conservation and Development.* London and Hanover: University Press of New England.

Thomas, H. and C. Logan, (1982). *Mondragon: An Economic Analysis.* London: Allen & Unwin.

Tiger, L. and R. Fox (1971). *The Imperial Animal.* New York: Holton, Rinehart & Winston.

Tilgher, A. (1930). *Work, What it has Meant to Men Through the Ages (Homo faber).* New York: Harcourt, Brace & Co.

Tocqueville, A. de (1835). *Democracy in America.* 2 vols. New York: Vintage Books, 1945.

Trist, E., G. Higgin, H. Murray, and A. Pollack (1963) *Organizational Choice*. London: Tavistock.

Tullock, G. (1967). The general irrelevance of the general impossibility theorem. *Quarterly Journal of Economics* **81**: 256—270.

Tuomela, R. (1977). *Human Action and its Explanation*. Dordrecht and Boston: Reidel.

Tuomela, R. (1984). *A Theory of Social Action*. Dordrecht and Boston: Reidel.

Tuomela, R. (1985). *Science, Action, and Reality*. Dordrecht and Boston: Reidel.

Turkish Comisso, E. (1979). *Workers' Control Under Plan and Market*. New Haven and London: Yale University Press.

Tversky, A. (1969). The intransitivity of preference. *Psychological Review* **76**: 31—48.

Vanek, J., Ed. (1975). *Self Management, Economic Liberation of Man*. Harmondsworth: Penguin.

Veblen, T. (1899). *The Theory of the Leisure Class*. New York: The Modern Library, 1934.

Vickrey, W. (1975). Private philanthropy and public finance. In Phelps, Ed., pp. 149—169.

Vining, D. R. (1986). Social versus reproductive success: The central theoretical problem of human sociobiology. *Behavioral and Brain Sciences* **9**: 167—216.

Von Neumann, J. and O. Morgenstern (1947). *Theory of Games and Economic Behavior*, 2nd ed. Princeton: Princeton University Press.

Von Wright, G. H. (1963). *The Varieties of Goodness*. London: Routledge & Kegan Paul.

Von Wright, G. H. (1968). *An Essay in Deontic Logic and the General Theory of Action*. Helsinki: Societas Philosophica Fennica.

Von Wright, G. H. (1971). *Explanation and Understanding*. Ithaca, N.Y.: Cornell University Press.

Waddington, C. H. (1960). *The Ethical Animal*. London: Allen and Unwin.

Walzer, M. (1977). *Just and Unjust Wars*. New York: Basic Books.

Wasserstrom, R. (1985). War, nuclear war, and nuclear deterrence: some conceptual and moral issues. *Ethics 95*: 424—444.

WCED (1987). *Our Common Future* [Brundtland Report]. Oxford: Oxford University Press.

Weber, M. (1904—1905). *Die protestantische Ethik und der Geist des Kapitalismus*. In *Gesammelte Aufsätze zur Religionssoziologie* I, pp. 17—206. Tübingen: J.C.B. Mohr, 1988.

Weber, M. (1921). Politics as a vocation. In *From Max Weber: Essays in Sociology*, pp. 77—128, H. H. Gerth and C. Wright Mills, Eds. New York: Oxford University Press, 1946.

Weinberger, O. (1985). Freedom, range for action, and the ontology of norms. *Synthèse* **65**: 307—324.

Weiss, R. F., W. Buchanan, L. Altstatt and J. P. Lombardo (1971). Altruism is rewarding. *Science* **171**: 1262—1263.

Wellman, C. (1971). *Challenge and Response. Justification in Ethics*. Carbondale IL: Southern Illinois University Press.

Wells, H. G. (1916). *Mr Britling Sees it Through*. New York: Macmillan.

Westin, A. (1986). Professional and ethical dissent. *Technology in Society* **8**: 335—339.

Williams, B. (1972). *Morality: An Introduction to Ethics.* New York: Harper & Row.

Williams, B. (1973). A Critique of Utilitarianism. In Smart and Williams, pp. 75—150.

Williams, B. (1986). *Ethics and the Limits of Philosophy.* Cambridge, MA: Harvard University Press.

Wilson, E. O. (1975). *Sociobiology: The Modern Synthesis.* Cambridge, MA: Belknap Press.

Wilson, E. O. (1978). *On Human Nature.* Cambridge, MA: Harvard University Press.

Wilson, T. C. (1985). Urbanism and tolerance: A test of some hypotheses drawn from Wirth and Stouffer. *Amer. Sociol. Rev.* **50**: 117—123.

Winch, P. (1958). *The Idea of a Social Science.* London: Routledge & Kegan Paul.

Wren, T. E. (1982). Social learning theory, self-regulation, and morality. *Ethics* **92**: 409—424.

Zahn-Waxler, C., E. M. Cummings, and R. Ianotti, Eds. (1986). *Altruism and Aggression: Biological and Social Origins.* Cambridge: Cambridge University Press.

Zayan, R. (1984). *Recherches d'éthologie appliquée sur la poule domestique.* Louvain: Université Catholique de Louvain.

Zayan, R., Ed. (1985). *Social Space for Domestic Animals.* Dordrecht: Martinus Nijhoff for the Commission of European Communities.

Zwerdlin, D. (1980). *Workplace Democracy.* New York: Harper & Row.

INDEX OF NAMES

Ackerman, R. 163, 182, 186, 190, 225
Ackoff, R. 56, 343
Afriat, S. 54
Agassi, J. xv, 262, 343
Aigner, T. 26
Albert, H. 75, 200, 351
Alchourrón, C. E. xv, 97, 298
Alexander, R. D. 121
Allais, M. 83
Almeder, R. 277
Altstatt, L. 123
Anand, P. 83
Aquinas, T. 2, 231, 232
Aristotle 2, 41, 44, 64, 69, 72, 75, 135, 178, 179, 213, 290, 333, 370, 375
Aronfred, J. 122, 135
Arrow, K. 70, 280, 337
Assisi, Francis of 199
Augustine 57
Austin, J. 237
Axelrod, R. 121, 141, 149
Ayer, A. J. 73, 207

Baier, K. 207
Balster, R. L. 26
Bandura, A. 122, 135
Batson, C. D. 125
Beer, S. 54
Bellon, B. 77, 368
Bentham, J. 2, 42, 72, 87, 107, 213, 236—237, 241, 269, 290
Berger, R. E. 309
Bergson, H. 72, 120, 199, 282
Bernoulli, D. 87
Bidney, D. 65
Black, M. 304
Blackwell, K. 260
Blitz, D. xv
Block, J. 30
Bodde, D. L. 261
Bonner, J. 337
Borovoy, A. A. 226
Bouglé, C. 70
Brandt, R. B. 201, 209, 237
Broad, C. D. 201, 293
Bronowski, J. 256
Brown, L. R. 357, 363, 387
Bruera, J. J. 271
Buber Agassi, J. 333, 350
Buchanan, J. 70, 186, 226
Buchanan, W. 123
Bulygin, E. xv, 97, 298
Brecht, A. 65
Bunge, A. xv
Bunge, C. O. 286
Bunge, M. xiv, 14, 24, 25, 37, 124, 135, 250, 252, 301, 321, 325, 347, 349
Burki, S. 34
Butler, J. 122

Cajal, S. R. y 256
Calvin, J. 134, 153, 175
Calvocoressi, P. 282
Camus, A. 218
Carr, E. H. 388
Castañeda, H.-N. xv
Cerf, J. A. 148
Chagnon, N. 146, 151
Chazan, B. 138
Cohn, P. 268
Coke, J. S. 125
Collins, J. 190
Comte, A. 253
Condorcet, A.-N. de 280
Confucius 213
Constant, B. 169
Cragg, W. 268
Cummings, E. M. 30

Dafoe, J. 137

Dahl, R. A. 225, 350
Darwin, C. R. 121, 122—123, 221, 239, 257
Dawkins, R. 134
Denenberg, V. 123
Dewey, J. 112
Dickens, C. 123
Diderot, D. 231, 232
Dix, T. 137
Durkheim, E. 70, 72, 130, 131, 213, 290, 347
Dworkin, R. 186, 275

Eatwell, J. 258
Ebling, F. J. G. 120
Eccles, J. C. 253
Eichner, A. S. 86
Einstein, A. 169, 312, 314
Emery, F. E. 56
Engelhardt, H. T. 268
Engels, F. 130, 131, 232, 253
Epicurus 229, 252
Evarts, E. V. 28

Feinberg, J. 70, 274
Ferrater-Mora, J. M. xv, 268
Feuerbach, L. 333
Field, H. H. 358
Fishkin, J. S. 206
Forman, F. xv, 227
Forsyth, D. R. 309
Fox, R. 146
Franco, F. 169, 199
Franklin, B. 382
Freud, S. 126, 141, 146, 188
Friedman, M. 188, 276
Friedrich Wilhelm II 210
Frondizi, R. xv, 64, 230

Gale, S. 330
Gandhi, M. 167, 380
Garzón-Valdés, E. xv, 161, 224, 379
Gasparski, W. 327
Gauthier, D. 132, 218, 219, 226, 228, 229, 240, 276
Geach, P. 14
Ghai, D. P. 34
Gilbert, J. O. 33
Gilligan, C. 122, 136
Gilligan, J. 218
Giudice, S. 136
Glover, J. 223
Goslan, D. A. 122
Graham, L. R. 78
Greene, D. 137
Grotius, H. 231, 232
Grousec, J. E. 137

Hagen, O. 83
Hahn, F. 86
Halal, W. E. 350, 369
Hamilton, W. D. 121
Hampshire, S. 115, 238
Handy, R. 19
Hardin, R. 239
Hare, R. M. 204
Harman, G. 308
Harris, J. 236
Harris, M. 34, 121, 312
Harsanyi, J. 235, 236, 237, 290
Hart, H. L. A. 272, 325
Hartmann, N. 64, 290
Harvey, P. H. 123, 174
Hayek, F. A. 188, 224, 342
Hebb, D. O. 252, 332
Hegel, G. W. F. 65, 178, 192, 213, 219, 232, 253, 333
Helvétius, C.-A. 42
Heraclitus 254
Herbst, P. G. 350
Herder, J. G. v. 232
Herskovits, M. 70
Hesburgh, T. M. 179, 381
Hippocrates 229—230
Hitler, A. 219
Hobbes, T. 2, 72, 103, 122, 125, 131, 132, 134, 139, 145, 146, 147, 220, 221, 226—227, 229, 231, 232, 290
Homans, G. 69
Honoré, A. M. 325
Hörz, H. E. 131
Hospers, J. 183, 226
Hull, C. 19, 26
Humber, J. 277

Hume, D. 2, 72, 73, 143, 197, 207, 213, 244, 269, 304—306
Huxley, A. 365
Huxley, T. H. 121

Ianotti, R. 30
Ingenieros, J. 254, 297

Jantsch, E. 330
Jefferson, T. 41, 375, 383
Juan Carlos I 169

Kálnay, A. J. xv
Kamin, L. J. 134, 145
Kanger, S. 96, 297
Kant, I. 41, 66, 70, 107, 132, 134, 136, 144, 153, 169, 172, 182, 197, 204, 210, 213, 231, 232, 233—235, 242, 254, 290, 292—293, 295
Kaufmann, W. 188
Kautsky, K. 123
Kelsen, H. 270—271
Keynes, J. M. 387
Kitcher, P. 141, 145
Knoke, D. 214
Kohlberg, L. 122, 132, 135
Kolnai, A. 209, 292
Korsgaard, C. M. 75
Koshland, D. D. Jr 24
Koslowski, P. 276
Kotarbinski, T. 327
Kraft, V. 74
Kropotkin, P. 103, 123, 147, 148, 376, 388
Kurtz, P. 201

Landsell, H. 269
Lee, V. 256
Leibniz, G. W. 159, 254
Lemmon, E. J. 296
Lenin, V. I. 131, 213, 237, 279
Leonard, H. S. 297
Lepper, M. R. 137
Lévy-Bruhl, L. 70
Lewis, C. I. 71, 74, 207, 210, 245, 290
Lewis, C. S. 216
Lewis, W. A. 152
Light, R. J. 33
Locke, J. 183, 309
Lombardo, J. P. 123
Lorenz, K. 146
Luce, R. D. 42
Lumsden, J. C. 145
Luther, M. 175

Machiavelli, N. 237, 278
MacIntyre, A. 305
Mackie, J. L. 67, 129, 305
Malinowski, B. 34
March, J. G. 54, 107
Marković, M. xv, 334, 350
Martínez-Selva, J. M. 180
Marx, K. 2, 65, 72, 232, 237, 257, 322, 333, 347
Maslow, A. 38
Maynard Smith, J. 148
McHale, J. 34
McHale, M. 34
Mencken, H. L. 282
Mendel, G. 257
Mill, J. S. 2, 42, 70, 224, 236—237, 241, 364, 370
Miller, D. 190, 214, 350
Milner, P. 26
Miró-Quesada, F. 297
Moessinger, P. xv, 122, 140
Monod, J. 256
Montesquieu, C. de S. 131, 231
Moore, B. 388
Moore, G. E. 2, 13, 23, 37, 42, 108, 204, 290
Moore Lappé, F. 190
Morgenstern, O. 42, 83
Mosteller, F. 33, 240
Muller, E. N. 190
Mosterín, J. 325
Murdock, G. P. 129
Mussolini, B. 146, 184, 282
Myrdal, G. 77

Nader, R. 260
Neurath, O. 342
Neyrinck, J. 261
Nielsen, K. 186, 190

Newton, I. 257
Nienhaus, V. 369
Nietzsche, F. 22, 60, 67, 169, 219—220
Nikolai, G. 169
Nino, C. 186, 308
Niosi, J. 77, 369
Nisbett, R. 137
Nozick, R. 176, 224
Nriagu, J. O. 357

Olds, J. 26
Olson, M. 338, 347
Olweus, D. 30
Opp, K.-D. 140
Oppenheimer, J. R. 259
Ortega y Gasset, J. 64
Otis, L. S. 148

Pacyna, J. M. 357
Palme, O. 279
Panksepp, J. 121
Pareto, V. 143, 229—230
Parfit, D. 302
Parra-Luna, F. 19
Paul 172
Pericles 279
Perry, R. B. 67, 290
Petkas, R. 260
Piaget, J. 108, 122, 135
Plato 2, 11, 57, 64, 69, 131, 137
Plotinus 27, 69
Pollock, J. L. 208
Pollock, L. 250
Popper, K. R. 75, 230
Postel, S. 387
Potter, N. T. 235
Priestley, J. 42
Protagoras 67
Pzsczolowski, W. 327
Puelles, L. xv, 83

Quintanilla, M. A. xv

Radke-Yarrow, M. 30
Raiffa, H. 42
Rand, A. 221
Ratner, L. G. 238
Rawls, J. 37, 182, 187, 192, 213, 224, 227, 240, 251, 302
Regan, D. 239
Rescher, N. 167, 213, 223, 239, 263
Rogers, C. 38
Robinson, D. N. 253
Robinson, J. 77, 144, 258
Rodríguez-Campoamor, H. xv
Rohatyn, F. 216
Ross, W. D. 108, 204, 207
Rousseau, J.-J. 145, 226—227, 231
Russell, B. xiii, 67, 72, 161, 169, 179, 186
Russell, E. S. 134
Rutherford, E. 257

Scanlon, T. C. 228
Schaub, J. H. 260
Scheffler, S. 212
Scheler, M. 64, 108, 207
Schlick, M. 74, 87, 262
Schweikart, D. 350
Scitovsky, T. 333
Scrooge 223
Seebass, G. 319
Selucky, R. 350
Sen, A. 43, 87, 223, 240, 258, 276, 290, 337
Settle, T. xv
Shaw, G. B. 17
Shinoda, Y. 28
Sidgwick, H. 42, 108, 237
Simon, H. A. 54, 107
Simpson, G. G. 112
Singer, P. 366
Sivard, R. L. 48, 77, 158, 390
Sjoberg, G. 256
Smart, J. J. C. 236, 238
Smart, R. N. 230
Smith, A. 77, 122, 370
Snoeyenbos, M. 277
Socrates 2, 137
Skinner, B. F. 71, 72
Snow, C. P. 256
Spicker, S. F. 268
Spinoza, B. 2, 27, 72, 122, 179, 213, 220, 221, 253, 254, 290, 309

Spurgeon, D. 171
Stent, G. 120
Stern, A. xv, 64, 69, 233, 388
Stevenson, C. L. 73, 208
Stevenson, R. L. 120, 144
Stojanović, S. 369
Stone, C. 273
Streeten, P. 34
Suárez, F. 312

Talls, K. 123, 174
Taylor, M. 376
Teller, E. 259
Teresa, Mother 252
Thatcher, R. W. 136
Thibodeau, F. R. 358
Thoreau, H. D. 380
Tiger, L. 146
Tilgher, A. 175
Timmons, M. 235
Tocqueville, A. de 172, 180, 214
Tolstoy, L. 251
Trasymachus 131
Tuomela, R. 302, 319, 321
Turkish Comisso, E. 350
Tutu, D. 199

Vanek, J. 334, 350
Veblen, O. 48, 173, 307, 332
Vickrey, W. 171
Vico, G. B. 65
Vierkandt, A. 70
Vining, D. R. 145
Voltaire 105
von Neumann, J. 42, 83
von Wright, G. H. 302
Vonnegut, K. 334

Waddington, C. H. 314
Walker, R. A. 136
Walzer, M. 282
Weber, M. 67, 77, 280, 309, 321
Weinberger, O. 296
Weingartner, P. xv
Weiss, R. F. 123
Wells, H. G. 130, 390
Westermarck, E. 70
Westin, A. 260
Wilke, U. 131
Williams, B. 14, 85–86
Wilson, E. O. 121, 145
Wilson, T. C. 130
Winch, P. 204, 205
Wise, S. P. 28
Wittgenstein, L. 285

Zahn-Waxler, C. 30
Zayan, R. 46
Zeno the Stoic 144
Zwerdlin, D. 334, 350

INDEX OF SUBJECTS

Abortion 164—165, 267
Absolutism
  anthropological 64—67
  axiological 13, 64—67
  ethical 204—205
    see also Deontology, Kant
  ontological 64—65
Action 264—284
  collective 335—353
  individual 319—334
  morally right 5, 106—108
Action theory 5, 7, 319—353
  see also Praxiology
Ad hominem argument 303
Agathonism 5, Chapters 4—6, 241—242, 394—395, 398
Agent 108, 323
  see also Action
Aggregation 46, 239—240
Aggression 147—148
Aid, foreign 171
Alienation 333
Altruism 123, 167—169, 231—242
  reciprocal 148
  see also Cooperation
Analysis, philosophical 2, 3, 285—286
Animal rights 101, 269—270, 365—366
Argument, moral 245—248, 299—303
Arms race 179, 381
Ars moriendi 166
Ars vivendi 166
Artifact 325
Asilomar convention 270
Athletism 268
Authoritarianism, ethical 295—296, 339
Authority 379
Autonomy, moral 215, 245, 254
Axiology 1, 5—6, 61—90, 290—291

Behavior
  antisocial 29—31
  constraints on 93
Bioethics 266—270, 362—367
Biologism, ethical 121
Biology 119—122
Biovalue 19—23
Birth control 200, 363—364
Blame 110, 215—216
Buddhism 166, 322
Business ethics 276—278

Calculation in moral matters 222—223
Calculus of value xiv, 88—89
Capitalism 309, 368—369, 387
Categorical imperative 210, 235
Categorical principle 300
Centralization 338—339
Change in morals 133—157
Charity 170—171
Children, rights of 163
Choice 28—29, 70, 82
Christianity 199—200, 210, 253
Circle, moral
  vicious 170
  virtuous 170
Code
  legal 94
  moral 94, 116, 128—132
Cognition 71
Cognitive-developmental theory 135—138
Competition 146—148, 345—347, 369—370
Compromise 54, 280
Computer 155—156
Conditional 295—296
Conflict
  between freedoms 225—226

  moral 143—144
  of norms 111—112
  of values 111—112
Consciousness 72
  moral 125—127
Consequentialism 107, 209—212
Consequentialist principle 217
Conservatism 216
Consistency, moral 175
Contractualism 219, 226—229
Cooperation 148—151, 337, 343—349
  competitive 149—150, 345—346, 347—349, 370—371
Cooperative firm 369—370, 377
Corrigibility 5
Cosmopolis 382
Counterfactual 321
Credit 110
Crime
  legal 274—275
  moral 274
Crisis, moral 154—156
Cultural philosophy 383—387
Culture 383—387

Death penalty 165, 252
Decentralization 339—340, 377
Decision theory 42—44, 291—292
Déclaration des droits de l'homme et du citoyen 41, 154, 232, 275
Declaration of Independence, American 41, 181, 232
Declaration of The Hague 362
Declarative 295—297
Democracy 339
  integral 191—192, 339, 370—371, 378—380, 395—396
  political 152, 190
Deontologism 107, 209—211, 233—235
  see also Kant
Desirable 69
Determinism 252—253
Dialectics 347
Dignity 171
Distribution of goods and burdens 188
Dogmatism 208—209, 234—235, 297
  see also Absolutism
Drug addiction 273, 365
Duty 4, 95—106
  doing one's 176
  evolution of 152—154
  kinds of 100
  legal 96
  moral 96—106
  natural 119
  negative 105
  positive 105
  private 160—164
  professional 160—161
  public 160—161

Economic philosophy 367—374
Economics 76—77, 258, 367—368
Ecosociodevelopment 358—362
Education 385—386
  moral 153—154
Efficiency 313, 330—331
Egalitarianism 225, 236
  qualified 182
  radical 177, 181
Egoism 167—169, 218—231
  rational 220—223
Emotion 71
  moral 122—128
Emotivism, ethical 73, 207—208
Empathy 125—126
Endomorality 255—256, 258—259
Exomorality 256, 259—263
Endovalue 76
Exovalue 76
Engineering, genetic 270, 365
Enlightenment 219
Entitlement 176
Entropy 325
Environment, management of 356—362
Environmental
  degradation 179
  protection 179, 356—362
Environmentalism
  romantic 357
  scientific 357—358
Epistemology 248—251
Equality 180—184, 225

Erogatory 176
Ethics 1, 197—315
  applied 243—244
  philosophical 287—290, 295
  Protestant 309
  scientific 287—290, 295
Euthanasia 165—166, 267—268
Evolution 144
  of mores and morals 151—154
Existentialism 33
Experiment
  ethical 140
  on animals 268—269
  social 340
Explanation 257

Fact 297
  moral 206, 286—287
Fact-value gap 12, 71—73, 249, 301
Fairness 108, 187
Fallacy, naturalistic 23, 37, 72
Fallibilism 313
Fascism 131, 146
Feeling, moral 122—123, 137
  see also Emotion, moral
Fiction 302
Fitness 20—21
Frame, general 292
Forgivingness 171—172
Fraternity 172, 180
Freedom 28—29, 162—163, 184—186, 252—254, 313—314, 321, 342, 378
  active 184
  kinds of 185
  passive 184
Frugality 173

Game theory 42, 141, 148—150
Gene, selfish 134
Geneticism 145
Golden Rule 103, 161, 206—207, 293, 309
Good 5—6, 14, 36, 42, 69, 82, 234
  basic 37
  public 376—377
Good will 172
Government 351
  see also State

Happiness 42—47
  reasonable 48—51
Health 21—23
  care 364
Hedonism 48, 220
Helpfulness 172
Heteronomy 254
Hippocratic Oath 283
History of morals 139—140
Holism
  axiological 70
  ethical 155, 212—216
  sociological 392
Homosexuality 174
Honesty 172—173
Humanism 41, 218
Humility 173

Idealism 71, 321
Ideals 55—58
Imperative 295—297
Inaction 325
Income, just 373
Indicator 82
Individualism
  axiological 70
  ethical 147, 155, 212—216
  sociological 392
Inference rule, axiological 301—302
Intention 107, 323—324
Interest
  private 335—336
  public 335—336
Intuitionism 108
  ethical 208—209
Invisible hand 229
Is-ought gap 301, 304—307

Junk culture 385—387
Justice 50, 187—190
  distributive 240
  international 373—374
  social 50, 188—190, 372—373
Justification 247—248

  moral 335—336
  of actions 327—328

Kaldor—Hicks criterion 261
Kantianism 6, 103, 209, 251

Law 227—228, 270—276
Law and Order Dilemma 275
Leadership 233
Libertarianism 223—226, 275—277, 364—365
Liberty *see* Freedom
Life, right to 162—166
Logic 244—248, 320
  deontic 246—247, 296—297
  modal 296—297

Machiavellism 63, 212
Management 349—352
  participatory 350
Marxism 277
Mass media 386
Materialism 254
  eliminative 252
  emergentist 13, 252—254
  physicalist 13, 71, 254
Maximization 54, 85
  of utility 239—240
Meaning 73
Means-end relation 62—63, 78—79, 84—86, 211—212, 239, 262—263, 330—332
Meliorism 313—314
Merit 186—177
Meta-axiology 1
Metaethics 1, 285—315
Metanorm 116—118, 235
Metaphysics 13
  see also Ontology
Metapraxiology 1
Meta-rule 292—293
Misery 49—50
Models of man 197
Modus nolens 301, 305
Modus volens 301, 305
Monism, ethical 201—204
Moral sphere
  private 159—162
  professional
  public 159—162
Morale 132
Morals 11, 93—193
  biological roots 119—122
  psychological roots 122—128
  social roots 128—132
Murder 165—166, 177—178
  see also War

Nativism 134
Natural law 231—233, 308
Naturalism 71
Nature, human 143—145, 395
Nature, state of 227
Nazism 274, 347
Need 4, 35
  basic 35
  primary 35
  secondary 35
Nihilism
  axiological 33
  ethical 218—220
Nomoethics 270—276
Norm 4, 116
  origin of 139—142

Objectivism, ethical 205—207
Objectivity 108
Ontology 251—255
  see also Metaphysics
Optimization 85
Organization 338—340
Overpopulation 362—364

Pareto optimality 261
Participation 192, 279, 378
  see also Democracy
Passion 244
  see also Emotion
Paternalism, legal, 275, 379
Patriotism 281—282
Perestroika 369
Perfectibility 57—58
  see also Meliorism
Perfection 57—58

Perfectionism 115
Permissiveness 163
Person 273, 302—303
Pessimism 143
Phenomenology 33
Philosophy
  cultural 383—387
  economic 367—374
  environmental 356—362
  legal 270—276
  linguistic 285
  moral 93—317
  political 374—383
  practical 7, 243
  social 354—389
Physiology, moral 125, 127
Plan 330—332, 341—343
Platonism 33
Pluralism, ethical 201—204
Policy 340—343
Political ethics 278—283
Political philosophy 374—383
Positivism 73
  legal 6, 270—271, 309
Poverty 50, 158, 377
Pragmatism 256
Praxiology 1, 327, 334
  see also Action theory
Praxis 264—284
Predicate 169
Preference 82
Price 18, 19
Prisoner's Dilemma 141, 150
  See also Game theory
Privilege 191
Probability 42—43
Problem
  global 390—392
  moral 111—115
  orders of magnitude of 114
  social 112—113
  typology of 113
Procreation, right to 164
Progress, moral 152—156
Project, national 388—389
Proof in ethics 299—300
Property
  emergent 15—16
  relational 15
  secondary 15
Property rights 182—183, 224, 232, 374
Propositionalization of imperatives 295—298
Prosperity 50
Pseudovalue 68
Psychoanalysis 141—142, 219
Psychovalue 23—29
Psychology 122—128, 266—267

Quid pro quo principle 103

Rationalism 75, 94, 132
Rationality 245, 325—326, 367
Realism 5, 249—251, 321
Reasoning, moral 126—128, 244—248, 299—303
Reference 286
Reform, global 388
Reformism
  ethical 216—217
  social 355
Relativism
  axiological 65, 143
  cultural 65
  ethical 143, 204—205
Religion 142, 199—201
Resources
  non-renewable 158, 359
  renewable 359
Responsibility 109—111
  collective 110
Revolution
  American 152
  French 152, 184, 210
  industrial 356
  social 355—356, 380
Right 4, 95—106
  kinds of 152—154
  legal 96
  moral 96—106
  natural 95, 119
  negative 105
  positive 105

Risk 78—79
Rule 294—299
  see also Norm

Satisfaction 54
Science
  and ethics 255—258
  and values 75—78
  social 76—78
Secularism 201
Security 381
Selfishness 221—223
  see also Egoism
Self-management 190—191
  see also Democracy
Self-restraint 173—174
Selftuism 5, 104
  see also Agathonism
Semantics 264
Sin 166—170
  Christian 167
Sincerity 174—175
Situationalism 65
Size of an organization 350—351
Social choice theory 70, 337
Social-learning theory 135—138
Social system 29—33
Socialism 182, 333, 368—369, 387
Sociality 95
Socialization 137
Society
  good 51
  just 50
  miserable 49—50
  poor 50
  prosperous 50
  welfare 51
Sociobiology 121, 140—141, 145, 149
Sociovalue 29—33
Sociologism, ethical 131—132, 143
Sociotechnology 331
Solidarity 172, 347
Soul 164, 252
Sovereignty, national 382
State 51, 146, 224—225, 374—383
  relief 51, 369
  welfare 51, 369, 375
State space 45—46, 307, 329—331
Stoicism 47, 72, 166, 169, 209, 231—232
Strategy *see* Plan
Subjectivism, ethical 205—207
Suicide 165—166
Summum bonum 59—60
Supererogatory 176
Survival, human 59—60, 117, 342, 389, 393
Syllogism, practical 302
  see also Inference Rule, axiological
System
  of freedoms 185
  of norms 298—299
  of virtues and sins 170
Systemic approach 310, 330, 356
Systemism
  axiological 71
  ethical 213—216
Systems theory 329

Technoclasm 261—262
Technocracy 378
Technodemocracy 340, 378—379, 397
Technolatry 261
Technology 78—79
  and ethics 258—263, 327, 334
Terrorism 177
Test
  of ethical theories 249—250, 311—314
  of moral norms 249
Testability 5
Theology 33, 251—252, 290—291
Theory, ethical 197—242
Tolerance 280—281
Truth 256, 280, 321
  moral 306—308

Universal Charter of Human Rights and Duties 395—398
Universal Declaration of Human Rights 154
Undererogatory 176
United States of the World 351, 381—382, 389, 397

Universality 5, 293
Unselfishness 108
Utilitarianism 17, 42—44, 84—85, 107, 112, 202, 211, 236—241
  act 237—238
  altruistic 236—241
  egoistic 236—237
  negative 229—230
  rule 237—238
Utility 18, 42—44, 86—88
  Principle of 43, 292—293

Value 4, 11—90
  aesthetic 74—75
  basic 81—82
  biological 12, 16
    see also Biovalue
  categories 38—40
  change 55
  conflict 26—27, 53—55
  de facto 66
  de jure 66
  extrinsic 62—63
  individual 69—71
  instrumental 62—63
  intrinsic 62—63
  judgment 13—14, 73—75, 301
  kinds of 16—18
  objective 67—69, 83—86
  physiological 20
  predicate 13—16
  primary 36
  psychological 12, 16
    see also Psychovalue
  qualitative 80
  quantitative 80—82
  quaternary 36
  secondary 36
  selective 20
  social 12–16, 69–71
    see also Sociovalue
  subjective 67—69, 86—88
    see also Psychovalue, Utility
  system 24—25
  technological 78—79
  tertiary 36
  theory 61—91
    see also Axiology
Veil of ignorance 192, 251, 302
  see also Rawls, J.
Verstehen 321—322
Violence 150—151
  political 282
Virtue 166—176
  ability 167
  character 167
  Christian 166—167
Voting 280, 337

Want 4
  illegitimate 27
  legitimate 27, 35
War 154—155, 282—283, 380—381
Waste 79
Well-being 41—60
Welfare 41—60
  social 49—52
  society 51
  state 275
Wildlife 366—367
Will, free 28, 186, 252—253
Wisdom 328
Work 332—335
  morals 175—176, 333
Workmanship 332
World system 32
  see also United States of the World